

(19) World Intellectual Property Organization
International Bureau



(43) International Publication Date
20 February 2003 (20.02.2003)

PCT

(10) International Publication Number
WO 03/013227 A2

(51) International Patent Classification⁷: **A01H**

(21) International Application Number: **PCT/US02/25805**

(22) International Filing Date: **9 August 2002 (09.08.2002)**

(25) Filing Language: **English**

(26) Publication Language: **English**

(30) Priority Data:
60/310,847 9 August 2001 (09.08.2001) US
60/336,049 19 November 2001 (19.11.2001) US
60/338,692 11 December 2001 (11.12.2001) US
10/171,468 14 June 2002 (14.06.2002) US

(71) Applicant (for all designated States except US):
MENDEL BIOTECHNOLOGY, INC. [US/US]; 21375
Cabot Boulevard, Hayward, CA 94945 (US).

(72) Inventors; and

(75) Inventors/Applicants (for US only): **RATCLIFFE, Oliver** [GB/US]; 814 East 21st Street, Oakland, CA 94606 (US). **RIECHMANN, Jose, Luis** [ES/US]; 518 S. El Molino Avenue, #308, Pasadena, CA 91101 (US). **ADAM, Luc, J.** [CA/US]; 25800 Industrial Boulevard, Apt. L403, Hayward, CA 94545 (US). **DUBELL, Arnold, T.** [US/US]; 14857 Wake Avenue, San Leandro, CA 94578 (US). **HEARD, Jacqueline, E.** [US/US]; 810 Guilford Avenue, San Mateo, CA 94402 (US). **PILGRIM, Marsha, L.** [US/US]; 1368 Patrick Henry Drive, Phoenixville, PA 19460 (US). **JIANG, Cai-Zhong** [US/US]; 34495 Heathrow Terrace, Fremont, CA 94555 (US). **REUBER, T., Lynne** [US/US]; 1115 S. Grant Street, San Mateo, CA

94402 (US). **CREELMAN, Robert, A.** [US/US]; 2801 Jennifer Drive, Castro Valley, CA 94546 (US). **PINEDA, Omaira** [CO/US]; 4060 9th Place, Vero Beach, FL 32960 (US). **YU, Guo-Liang** [US/US]; 242 Gravett Drive, Berkeley, CA 94705-1531 (US). **BROUN, Pierre, E.** [FR/US]; 921 Sunnybrae Boulevard, San Mateo, CA 94402 (US).

(74) Agents: **WARD, Michael, R.** et al.; Morrison & Foerster LLP, 425 Market Street, San Francisco, CA 94105-2482 (US).

(81) Designated States (national): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, OM, PH, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZM, ZW.

(84) Designated States (regional): ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, SK, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:

— without international search report and to be republished upon receipt of that report

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

(54) Title: **YIELD-RELATED POLYNUCLEOTIDES AND POLYPEPTIDES IN PLANTS**

(57) Abstract: The invention relates to plant transcription factor polypeptides, polynucleotides that encode them, homologs from a variety of plant species, and methods of using the polynucleotides and polypeptides to produce transgenic plants having advantageous properties compared to a reference plant. Sequence information related to these polynucleotides and polypeptides can also be used in bioinformatic search methods and is also disclosed.

WO 03/013227 A2

YIELD-RELATED POLYNUCLEOTIDES AND POLYPEPTIDES IN PLANTS

This application claims the benefit of US Provisional Application No. 60/310,847, filed August 9, 2001, US Provisional Application No. 60/336,049, filed December 5, 2001, US Provisional Application No. 60/338,692, filed December 11, 2001, and US Non-provisional Application No. 10/171,468, filed June 14, 2002, the entire contents of which are hereby incorporated by reference.

FIELD OF THE INVENTION

This invention relates to the field of plant biology. More particularly, the present invention pertains to compositions and methods for phenotypically modifying a plant.

INTRODUCTION

A plant's traits, such as its biochemical, developmental, or phenotypic characteristics, may be controlled through a number of cellular processes. One important way to manipulate that control is through transcription factors - proteins that influence the expression of a particular gene or sets of genes. Transformed and transgenic plants that comprise cells having altered levels of at least one selected transcription factor, for example, possess advantageous or desirable traits. Strategies for manipulating traits by altering a plant cell's transcription factor content can therefore result in plants and crops with commercially valuable properties. Applicants have identified polynucleotides encoding transcription factors, developed numerous transgenic plants using these polynucleotides, and have analyzed the plants for a variety of important traits. In so doing, applicants have identified important polynucleotide and polypeptide sequences for producing commercially valuable plants and crops as well as the methods for making them and using them. Other aspects and embodiments of the invention are described below and can be derived from the teachings of this disclosure as a whole.

BACKGROUND OF THE INVENTION

Transcription factors (TFs) can modulate gene expression, either increasing or decreasing (inducing or repressing) the rate of transcription. This modulation results in differential levels of gene expression at various developmental stages, in different

tissues and cell types, and in response to different exogenous (e.g., environmental) and endogenous stimuli throughout the life cycle of the organism.

Because transcription factors are key controlling elements of biological pathways, altering the expression levels of one or more transcription factors can change entire biological pathways in an organism. For example, manipulation of the levels of selected transcription factors may result in increased expression of economically useful proteins or metabolic chemicals in plants or to improve other agriculturally relevant characteristics. Conversely, blocked or reduced expression of a transcription factor may reduce biosynthesis of unwanted compounds or remove an undesirable trait. Therefore, manipulating transcription factor levels in a plant offers tremendous potential in agricultural biotechnology for modifying a plant's traits.

The present invention provides novel transcription factors useful for modifying a plant's phenotype in desirable ways.

SUMMARY OF THE INVENTION

In a first aspect, the invention relates to a recombinant polynucleotide comprising a nucleotide sequence selected from the group consisting of: (a) a nucleotide sequence encoding a polypeptide comprising a polypeptide sequence selected from those of the Sequence Listing, SEQ ID NOs:2 to 2N, where $N = 2-561$, or those listed in Table 4, or a complementary nucleotide sequence thereof; (b) a nucleotide sequence encoding a polypeptide comprising a variant of a polypeptide of (a) having one or more, or between 1 and about 5, or between 1 and about 10, or between 1 and about 30, conservative amino acid substitutions; (c) a nucleotide sequence comprising a sequence selected from those of SEQ ID NOs:1 to $(2N - 1)$, where $N = 2-561$, or those included in Table 4, or a complementary nucleotide sequence thereof; (d) a nucleotide sequence comprising silent substitutions in a nucleotide sequence of (c); (e) a nucleotide sequence which hybridizes under stringent conditions over substantially the entire length of a nucleotide sequence of one or more of: (a), (b), (c), or (d); (f) a nucleotide sequence comprising at least 10 or 15, or at least about 20, or at least about 30 consecutive nucleotides of a sequence of any of (a)-(e), or at least 10 or 15, or at least about 20, or at least about 30 consecutive nucleotides outside of a region encoding a conserved domain of any of (a)-(e); (g) a

nucleotide sequence comprising a subsequence or fragment of any of (a)-(f), which subsequence or fragment encodes a polypeptide having a biological activity that modifies a plant's characteristic, functions as a transcription factor, or alters the level of transcription of a gene or transgene in a cell; (h) a nucleotide sequence having at least 31% sequence identity to a nucleotide sequence of any of (a)-(g); (i) a nucleotide sequence having at least 60%, or at least 70 %, or at least 80 %, or at least 90 %, or at least 95 % sequence identity to a nucleotide sequence of any of (a)-(g) or a 10 or 15 nucleotide, or at least about 20, or at least about 30 nucleotide region of a sequence of (a)-(g) that is outside of a region encoding a conserved domain; (j) a nucleotide sequence that encodes a polypeptide having at least 31% sequence identity to a polypeptide listed in Table 4, or the Sequence Listing; (k) a nucleotide sequence which encodes a polypeptide having at least 60%, or at least 70 %, or at least 80%, or at least 90 %, or at least 95 % sequence identity to a polypeptide listed in Table 4, or the Sequence Listing; and (l) a nucleotide sequence that encodes a conserved domain of a polypeptide having at least 85%, or at least 90%, or at least 95%, or at least 98% sequence identity to a conserved domain of a polypeptide listed in Table 4, or the Sequence Listing. The recombinant polynucleotide may further comprise a constitutive, inducible, or tissue-specific promoter operably linked to the nucleotide sequence. The invention also relates to compositions comprising at least two of the above-described polynucleotides.

In a second aspect, the invention comprises an isolated or recombinant polypeptide comprising a subsequence of at least about 10, or at least about 15, or at least about 20, or at least about 30 contiguous amino acids encoded by the recombinant or isolated polynucleotide described above, or comprising a subsequence of at least about 8, or at least about 12, or at least about 15, or at least about 20, or at least about 30 contiguous amino acids outside a conserved domain.

In a third aspect, the invention comprises an isolated or recombinant polynucleotide that encodes a polypeptide that is a paralog of the isolated polypeptide described above. In one aspect, the invention is an paralog which, when expressed in *Arabidopsis*, modifies a trait of the *Arabidopsis* plant.

In a fourth aspect, the invention comprises an isolated or recombinant polynucleotide that encodes a polypeptide that is an ortholog of the isolated polypeptide described above. In one aspect, the invention is an ortholog which, when expressed in *Arabidopsis*, modifies a trait of the *Arabidopsis* plant.

In a fifth aspect, the invention comprises an isolated polypeptide that is a paralog of the isolated polypeptide described above. In one aspect, the invention is an paralog which, when expressed in *Arabidopsis*, modifies a trait of the *Arabidopsis* plant.

In a sixth aspect, the invention comprises an isolated polypeptide that is an ortholog of the isolated polypeptide described above. In one aspect, the invention is an ortholog which, when expressed in *Arabidopsis*, modifies a trait of the *Arabidopsis* plant.

The present invention also encompasses transcription factor variants. A preferred transcription factor variant is one having at least 40% amino acid sequence identity, a more preferred transcription factor variant is one having at least 50% amino acid sequence identity and a most preferred transcription factor variant is one having at least 65% amino acid sequence identity to the transcription factor amino acid sequence SEQ ID NOs:2 to 2N, where N = 2-561, and which contains at least one functional or structural characteristic of the transcription factor amino acid sequence. Sequences having lesser degrees of identity but comparable biological activity are considered to be equivalents.

In another aspect, the invention is a transgenic plant comprising one or more of the above-described isolated or recombinant polynucleotides. In yet another aspect, the invention is a plant with altered expression levels of a polynucleotide described above or a plant with altered expression or activity levels of an above-described polypeptide. Further, the invention is a plant lacking a nucleotide sequence encoding a polypeptide described above or substantially lacking a polypeptide described above. The plant may be any plant, including, but not limited to, *Arabidopsis*, mustard, soybean, wheat, corn, potato, cotton, rice, oilseed rape, sunflower, alfalfa, sugarcane, turf, banana, blackberry, blueberry, strawberry,

raspberry, cantaloupe, carrot, cauliflower, coffee, cucumber, eggplant, grapes, honeydew, lettuce, mango, melon, onion, papaya, peas, peppers, pineapple, pumpkin, spinach, squash, sweet corn, tobacco, tomato, watermelon, rosaceous fruits, vegetable brassicas, and mint or other labiates. In yet another aspect, the inventions is an isolated plant material of a plant, including, but not limited to, plant tissue, fruit, seed, plant cell, embryo, protoplast, pollen, and the like. In yet another aspect, the invention is a transgenic plant tissue culture of regenerable cells, including, but not limited to, embryos, meristematic cells, microspores, protoplast, pollen, and the like.

In yet another aspect the invention is a transgenic plant comprising one or more of the above described polynucleotides wherein the encoded polypeptide is expressed and regulates transcription of a gene.

In a further aspect the invention provides a method of using the polynucleotide composition to breed a progeny plant from a transgenic plant including crossing plants, producing seeds from transgenic plants, and methods of breeding using transgenic plants, the method comprising transforming a plant with the polynucleotide composition to create a transgenic plant, crossing the transgenic plant with another plant, selecting seed, and growing the progeny plant from the seed.

In a further aspect, the invention provides a progeny plant derived from a parental plant wherein said progeny plant exhibits at least three fold greater messenger RNA levels than said parental plant, wherein the messenger RNA encodes a DNA-binding protein which is capable of binding to a DNA regulatory sequence and inducing expression of a plant trait gene, wherein the progeny plant is characterized by a change in the plant trait compared to said parental plant. In yet a further aspect, the progeny plant exhibits at least ten fold greater messenger RNA levels compared to said parental plant. In yet a further aspect, the progeny plant exhibits at least fifty fold greater messenger RNA levels compared to said parental plant.

In a further aspect, the invention relates to a cloning or expression vector comprising the isolated or recombinant polynucleotide described above or cells comprising the cloning or expression vector.

In yet a further aspect, the invention relates to a composition produced by incubating a polynucleotide of the invention with a nuclease, a restriction enzyme, a polymerase; a polymerase and a primer; a cloning vector, or with a cell.

Furthermore, the invention relates to a method for producing a plant having a modified trait. The method comprises altering the expression of an isolated or recombinant polynucleotide of the invention or altering the expression or activity of a polypeptide of the invention in a plant to produce a modified plant, and selecting the modified plant for a modified trait. In one aspect, the plant is a monocot plant. In another aspect, the plant is a dicot plant. In another aspect the recombinant polynucleotide is from a dicot plant and the plant is a monocot plant. In yet another aspect the recombinant polynucleotide is from a monocot plant and the plant is a dicot plant. In yet another aspect the recombinant polynucleotide is from a monocot plant and the plant is a monocot plant. In yet another aspect the recombinant polynucleotide is from a dicot plant and the plant is a dicot plant.

In another aspect, the invention is a transgenic plant comprising an isolated or recombinant polynucleotide encoding a polypeptide wherein the polypeptide is selected from the group consisting of SEQ ID NOs: 2 - 2N, where $N = 2-561$. In yet another aspect, the invention is a plant with altered expression levels of a polypeptide described above or a plant with altered expression or activity levels of an above-described polypeptide. Further, the invention is a plant lacking a polynucleotide sequence encoding a polypeptide described above or substantially lacking a polypeptide described above. The plant may be any plant, including, but not limited to, Arabidopsis, mustard, soybean, wheat, corn, potato, cotton, rice, oilseed rape, sunflower, alfalfa, sugarcane, turf, banana, blackberry, blueberry, strawberry, raspberry, cantaloupe, carrot, cauliflower, coffee, cucumber, eggplant, grapes, honeydew, lettuce, mango, melon, onion, papaya, peas, peppers, pineapple, pumpkin, spinach, squash, sweet corn, tobacco, tomato, watermelon, rosaceous fruits, vegetable brassicas, and mint or other labiates. In yet another aspect, the inventions is an isolated plant material of a plant, including, but not limited to, plant tissue, fruit, seed, plant cell, embryo, protoplast, pollen, and the like. In yet another aspect, the

invention is a transgenic plant tissue culture of regenerable cells, including, but not limited to, embryos, meristematic cells, microspores, protoplast, pollen, and the like.

In another aspect, the invention relates to a method of identifying a factor that is modulated by or interacts with a polypeptide encoded by a polynucleotide of the invention. The method comprises expressing a polypeptide encoded by the polynucleotide in a plant; and identifying at least one factor that is modulated by or interacts with the polypeptide. In one embodiment the method for identifying modulating or interacting factors is by detecting binding by the polypeptide to a promoter sequence, or by detecting interactions between an additional protein and the polypeptide in a yeast two hybrid system, or by detecting expression of a factor by hybridization to a microarray, subtractive hybridization, or differential display.

In yet another aspect, the invention is a method of identifying a molecule that modulates activity or expression of a polynucleotide or polypeptide of interest. The method comprises placing the molecule in contact with a plant comprising the polynucleotide or polypeptide encoded by the polynucleotide of the invention and monitoring one or more of the expression level of the polynucleotide in the plant, the expression level of the polypeptide in the plant, and modulation of an activity of the polypeptide in the plant.

In yet another aspect, the invention relates to an integrated system, computer or computer readable medium comprising one or more character strings corresponding to a polynucleotide of the invention, or to a polypeptide encoded by the polynucleotide. The integrated system, computer or computer readable medium may comprise a link between one or more sequence strings to a modified plant trait.

In yet another aspect, the invention is a method for identifying a sequence similar or homologous to one or more polynucleotides of the invention, or one or more polypeptides encoded by the polynucleotides. The method comprises providing a sequence database, and querying the sequence database with one or more target sequences corresponding to the one or more polynucleotides or to the one or more polypeptides to identify one or more sequence members of the database that display sequence similarity or homology to one or more of the one or more target sequences.

The method may further comprise of linking the one or more of the polynucleotides of the invention, or encoded polypeptides, to a modified plant phenotype.

BRIEF DESCRIPTION OF THE SEQUENCE LISTING, TABLES, AND FIGURE

The Sequence Listing provides exemplary polynucleotide and polypeptide sequences of the invention. The traits associated with the use of the sequences are included in the Examples.

Diskette1 is a read-only memory computer-readable diskette and contains a copy of the Sequence Listing in ASCII text format. The Sequence Listing is named "SEQLIST514442002041" and is 929 kilobytes in size. The copy of the Sequence Listing on the diskette is hereby incorporated by reference in its entirety.

Table 4 shows the polynucleotides and polypeptides identified by SEQ ID NO; Mendel Gene ID No.; conserved domain of the polypeptide; and if the polynucleotide was tested in a transgenic assay. The first column shows the polynucleotide SEQ ID NO; the second column shows the Mendel Gene ID No., GID; the third column shows the trait(s) resulting from the knock out or overexpression of the polynucleotide in the transgenic plant; the fourth column shows the category of the trait; the fifth column shows the transcription factor family to which the polynucleotide belongs; the sixth column ("Comment"), includes specific effects and utilities conferred by the polynucleotide of the first column; the seventh column shows the SEQ ID NO of the polypeptide encoded by the polynucleotide; and the eighth column shows the amino acid residue positions of the conserved domain in amino acid (AA) co-ordinates.

Table 5 lists a summary of orthologous and homologous sequences identified using BLAST (tblastx program). The first column shows the polynucleotide sequence identifier (SEQ ID NO), the second column shows the corresponding cDNA identifier (Gene ID), the third column shows the orthologous or homologous polynucleotide GenBank Accession Number (Test Sequence ID), the fourth column shows the

calculated probability value that the sequence identity is due to chance (Smallest Sum Probability), the fifth column shows the plant species from which the test sequence was isolated (Test Sequence Species), and the sixth column shows the orthologous or homologous test sequence GenBank annotation (Test Sequence GenBank Annotation).

Figure 1 shows a phylogenic tree of related plant families adapted from Daly et al. (2001 *Plant Physiology* 127:1328-1333).

Detailed Description of Exemplary Embodiments

In an important aspect, the present invention relates to polynucleotides and polypeptides, e.g. for modifying phenotypes of plants. Throughout this disclosure, various information sources are referred to and/or are specifically incorporated. The information sources include scientific journal articles, patent documents, textbooks, and World Wide Web browser-inactive page addresses, for example. While the reference to these information sources clearly indicates that they can be used by one of skill in the art, applicants specifically incorporate each and every one of the information sources cited herein, in their entirety, whether or not a specific mention of "incorporation by reference" is noted. The contents and teachings of each and every one of the information sources can be relied on and used to make and use embodiments of the invention.

It must be noted that as used herein and in the appended claims, the singular forms "a," "an," and "the" include plural reference unless the context clearly dictates otherwise. Thus, for example, a reference to "a plant" includes a plurality of such plants, and a reference to "a stress" is a reference to one or more stresses and equivalents thereof known to those skilled in the art, and so forth.

The polynucleotide sequences of the invention encode polypeptides that are members of well-known transcription factor families, including plant transcription factor families, as disclosed in Table 4. Generally, the transcription factors encoded by the present sequences are involved in cell differentiation and proliferation and the regulation of growth. Accordingly, one skilled in the art would recognize that by expressing the present sequences in a plant, one may change the expression of

autologous genes or induce the expression of introduced genes. By affecting the expression of similar autologous sequences in a plant that have the biological activity of the present sequences, or by introducing the present sequences into a plant, one may alter a plant's phenotype to one with improved traits. The sequences of the invention may also be used to transform a plant and introduce desirable traits not found in the wild-type cultivar or strain. Plants may then be selected for those that produce the most desirable degree of over- or underexpression of target genes of interest and coincident trait improvement.

The sequences of the present invention may be from any species, particularly plant species, in a naturally occurring form or from any source whether natural, synthetic, semi-synthetic or recombinant. The sequences of the invention may also include fragments of the present amino acid sequences. In this context, a "fragment" refers to a fragment of a polypeptide sequence which is at least 5 to about 15 amino acids in length, most preferably at least 14 amino acids, and which retain some biological activity of a transcription factor. Where "amino acid sequence" is recited to refer to an amino acid sequence of a naturally occurring protein molecule, "amino acid sequence" and like terms are not meant to limit the amino acid sequence to the complete native amino acid sequence associated with the recited protein molecule.

As one of ordinary skill in the art recognizes, transcription factors can be identified by the presence of a region or domain of structural similarity or identity to a specific consensus sequence or the presence of a specific consensus DNA-binding site or DNA-binding site motif (see, for example, Riechmann et al., (2000) *Science* 290: 2105-2110). The plant transcription factors may belong to one of the following transcription factor families: the AP2 (APETALA2) domain transcription factor family (Riechmann and Meyerowitz (1998) *Biol. Chem.* 379:633-646); the MYB transcription factor family (Martin and Paz-Ares, (1997) *Trends Genet.* 13:67-73); the MADS domain transcription factor family (Riechmann and Meyerowitz (1997) *Biol. Chem.* 378:1079-1101); the WRKY protein family (Ishiguro and Nakamura (1994) *Mol. Gen. Genet.* 244:563-571); the ankyrin-repeat protein family (Zhang et al. (1992) *Plant Cell* 4:1575-1588); the zinc finger protein (Z) family (Klug and Schwabe (1995) *FASEB J.* 9: 597-604); the homeobox (HB) protein family (Buerklin in *Guidebook to the Homeobox Genes*, Duboule (ed.) (1994) Oxford University Press);

the CAAT-element binding proteins (Forsburg and Guarente (1989) *Genes Dev.* 3:1166-1178); the squamosa promoter binding proteins (SPB) (Klein et al. (1996) *Mol. Gen. Genet.* 1996 250:7-16); the NAM protein family (Souer et al. (1996) *Cell* 85:159-170); the IAA/AUX proteins (Rouse et al. (1998) *Science* 279:1371-1373); the HLH/MYC protein family (Littlewood et al. (1994) *Prot. Profile* 1:639-709); the DNA-binding protein (DBP) family (Tucker et al. (1994) *EMBO J.* 13:2994-3002); the bZIP family of transcription factors (Foster et al. (1994) *FASEB J.* 8:192-200); the Box P-binding protein (the BPF-1) family (da Costa e Silva et al. (1993) *Plant J.* 4:125-135); the high mobility group (HMG) family (Bustin and Reeves (1996) *Prog. Nucl. Acids Res. Mol. Biol.* 54:35-100); the scarecrow (SCR) family (Di Laurenzio et al. (1996) *Cell* 86:423-433); the GF14 family (Wu et al. (1997) *Plant Physiol.* 114:1421-1431); the polycomb (PCOMB) family (Kennison (1995) *Annu. Rev. Genet.* 29:289-303); the teosinte branched (TEO) family (Luo et al. (1996) *Nature* 383:794-799); the ABI3 family (Giraudat et al. (1992) *Plant Cell* 4:1251-1261); the triple helix (TH) family (Dehesh et al. (1990) *Science* 250:1397-1399); the EIL family (Chao et al. (1997) *Cell* 89:1133-44); the AT-HOOK family (Reeves and Nissen (1990) *J. Biol. Chem.* 265:8573-8582); the S1FA family (Zhou et al. (1995) *Nucleic Acids Res.* 23:1165-1169); the bZIPT2 family (Lu and Ferl (1995) *Plant Physiol.* 109:723); the YABBY family (Bowman et al. (1999) *Development* 126:2387-96); the PAZ family (Bohmert et al. (1998) *EMBO J.* 17:170-80); a family of miscellaneous (MISC) transcription factors including the DPBF family (Kim et al. (1997) *Plant J.* 11:1237-1251) and the SPF1 family (Ishiguro and Nakamura (1994) *Mol. Gen. Genet.* 244:563-571); the golden (GLD) family (Hall et al. (1998) *Plant Cell* 10:925-936), the TUBBY family (Boggin et al. (1999) *Science* 286:2119-2125), the heat shock family (Wu C (1995) *Annu Rev Cell Dev Biol* 11:441-469), the ENBP family (Christiansen et al (1996) *Plant Mol Biol* 32:809-821), the RING-zinc family (Jensen et al. (1998) *FEBS letters* 436:283-287), the PDBP family (Janik et al *Virology.* (1989) 168:320-329), the PCF family (Cubas P, et al. *Plant J.* (1999) 18:215-22), the SRS (SHI-related) family (Fridborg et al *Plant Cell* (1999) 11:1019-1032), the CPP (cysteine-rich polycomb-like) family (Cvitanich et al *Proc. Natl. Acad. Sci. U S A.* (2000) 97:8163-8168), the ARF (auxin response factor) family (Ulmasov, et al. (1999) *Proc. Natl. Acad. Sci. USA* 96: 5844-5849), the SWI/SNF family (Collingwood et al *J. Mol. End.* 23:255-275), the ACBF family (Seguin et al (1997) *Plant Mol Biol.* 35:281-291), PCGL (CG-1 like) family (da Costa e Silva et al.

(1994) *Plant Mol Biol.* 25:921-924) the ARID family (Vazquez et al. (1999) *Development.* 126: 733-42), the Jumonji family, Balciunas et al (2000, *Trends Biochem Sci.* 25: 274-276), the bZIP-NIN family (Schauser et al (1999) *Nature* 402: 191-195), the E2F family Kaelin et al (1992) *Cell* 70: 351-364) and the GRF-like family (Knaap et al (2000) *Plant Physiol.* 122: 695-704). As indicated by any part of the list above and as known in the art, transcription factors have been sometimes categorized by class, family, and sub-family according to their structural content and consensus DNA-binding site motif, for example. Many of the classes and many of the families and sub-families are listed here. However, the inclusion of one sub-family and not another, or the inclusion of one family and not another, does not mean that the invention does not encompass polynucleotides or polypeptides of a certain family or sub-family. The list provided here is merely an example of the types of transcription factors and the knowledge available concerning the consensus sequences and consensus DNA-binding site motifs that help define them as known to those of skill in the art (each of the references noted above are specifically incorporated herein by reference). A transcription factor may include, but is not limited to, any polypeptide that can activate or repress transcription of a single gene or a number of genes. This polypeptide group includes, but is not limited to, DNA-binding proteins, DNA-binding protein binding proteins, protein kinases, protein phosphatases, GTP-binding proteins, and receptors, and the like.

In addition to methods for modifying a plant phenotype by employing one or more polynucleotides and polypeptides of the invention described herein, the polynucleotides and polypeptides of the invention have a variety of additional uses. These uses include their use in the recombinant production (i.e., expression) of proteins; as regulators of plant gene expression, as diagnostic probes for the presence of complementary or partially complementary nucleic acids (including for detection of natural coding nucleic acids); as substrates for further reactions, e.g., mutation reactions, PCR reactions, or the like; as substrates for cloning e.g., including digestion or ligation reactions; and for identifying exogenous or endogenous modulators of the transcription factors. A "polynucleotide" is a nucleic acid sequence comprising a plurality of polymerized nucleotides, e.g., at least about 15 consecutive polymerized nucleotides, optionally at least about 30 consecutive nucleotides, at least about 50 consecutive nucleotides. In many instances, a polynucleotide comprises a nucleotide

sequence encoding a polypeptide (or protein) or a domain or fragment thereof. Additionally, the polynucleotide may comprise a promoter, an intron, an enhancer region, a polyadenylation site, a translation initiation site, 5' or 3' untranslated regions, a reporter gene, a selectable marker, or the like. The polynucleotide can be single stranded or double stranded DNA or RNA. The polynucleotide optionally comprises modified bases or a modified backbone. The polynucleotide can be, e.g., genomic DNA or RNA, a transcript (such as an mRNA), a cDNA, a PCR product, a cloned DNA, a synthetic DNA or RNA, or the like. The polynucleotide can comprise a sequence in either sense or antisense orientations.

A "recombinant polynucleotide" is a polynucleotide that is not in its native state, e.g., the polynucleotide comprises a nucleotide sequence not found in nature, or the polynucleotide is in a context other than that in which it is naturally found, e.g., separated from nucleotide sequences with which it typically is in proximity in nature, or adjacent (or contiguous with) nucleotide sequences with which it typically is not in proximity. For example, the sequence at issue can be cloned into a vector, or otherwise recombined with one or more additional nucleic acid.

An "isolated polynucleotide" is a polynucleotide whether naturally occurring or recombinant, that is present outside the cell in which it is typically found in nature, whether purified or not. Optionally, an isolated polynucleotide is subject to one or more enrichment or purification procedures, e.g., cell lysis, extraction, centrifugation, precipitation, or the like.

A "polypeptide" is an amino acid sequence comprising a plurality of consecutive polymerized amino acid residues e.g., at least about 15 consecutive polymerized amino acid residues, optionally at least about 30 consecutive polymerized amino acid residues, at least about 50 consecutive polymerized amino acid residues. In many instances, a polypeptide comprises a polymerized amino acid residue sequence that is a transcription factor or a domain or portion or fragment thereof. Additionally, the polypeptide may comprise a localization domain, 2) an activation domain, 3) a repression domain, 4) an oligomerization domain or 5) a DNA-binding domain, or the like. The polypeptide optionally comprises modified

amino acid residues, naturally occurring amino acid residues not encoded by a codon, non-naturally occurring amino acid residues.

A "recombinant polypeptide" is a polypeptide produced by translation of a recombinant polynucleotide. A "synthetic polypeptide" is a polypeptide created by consecutive polymerization of isolated amino acid residues using methods well known in the art. An "isolated polypeptide," whether a naturally occurring or a recombinant polypeptide, is more enriched in (or out of) a cell than the polypeptide in its natural state in a wild type cell, e.g., more than about 5% enriched, more than about 10% enriched, or more than about 20%, or more than about 50%, or more, enriched, i.e., alternatively denoted: 105%, 110%, 120%, 150% or more, enriched relative to wild type standardized at 100%. Such an enrichment is not the result of a natural response of a wild type plant. Alternatively, or additionally, the isolated polypeptide is separated from other cellular components with which it is typically associated, e.g., by any of the various protein purification methods herein.

"Identity" or "similarity" refers to sequence similarity between two polynucleotide sequences or between two polypeptide sequences, with identity being a more strict comparison. The phrases "percent identity" and "% identity" refer to the percentage of sequence similarity found in a comparison of two or more polynucleotide sequences or two or more polypeptide sequences. Identity or similarity can be determined by comparing a position in each sequence that may be aligned for purposes of comparison. When a position in the compared sequence is occupied by the same nucleotide base or amino acid, then the molecules are identical at that position. A degree of similarity or identity between polynucleotide sequences is a function of the number of identical or matching nucleotides at positions shared by the polynucleotide sequences. A degree of identity of polypeptide sequences is a function of the number of identical amino acids at positions shared by the polypeptide sequences. A degree of homology or similarity of polypeptide sequences is a function of the number of amino acids, i.e., structurally related, at positions shared by the polypeptide sequences.

"Altered" nucleic acid sequences encoding polypeptide include those sequences with deletions, insertions, or substitutions of different nucleotides, resulting

in a polynucleotide encoding a polypeptide with at least one functional characteristic of the polypeptide. Included within this definition are polymorphisms that may or may not be readily detectable using a particular oligonucleotide probe of the polynucleotide encoding polypeptide, and improper or unexpected hybridization to allelic variants, with a locus other than the normal chromosomal locus for the polynucleotide sequence encoding polypeptide. The encoded polypeptide protein may also be "altered", and may contain deletions, insertions, or substitutions of amino acid residues that produce a silent change and result in a functionally equivalent polypeptide. Deliberate amino acid substitutions may be made on the basis of similarity in residue side chain chemistry, including, but not limited to, polarity, charge, solubility, hydrophobicity, hydrophilicity, and/or the amphipathic nature of the residues, as long as the biological activity of polypeptide is retained. For example, negatively charged amino acids may include aspartic acid and glutamic acid, positively charged amino acids may include lysine and arginine, and amino acids with uncharged polar head groups having similar hydrophilicity values may include leucine, isoleucine, and valine; glycine and alanine; asparagine and glutamine; serine and threonine; and phenylalanine and tyrosine. Alignments between different polypeptide sequences may be used to calculate "percentage sequence similarity".

The term "plant" includes whole plants, shoot vegetative organs/structures (*e.g.*, leaves, stems and tubers), roots, flowers and floral organs/structures (*e.g.*, bracts, sepals, petals, stamens, carpels, anthers and ovules), seed (including embryo, endosperm, and seed coat) and fruit (the mature ovary), plant tissue (*e.g.*, vascular tissue, ground tissue, and the like) and cells (*e.g.*, guard cells, egg cells, and the like), and progeny of same. The class of plants that can be used in the method of the invention is generally as broad as the class of higher and lower plants amenable to transformation techniques, including angiosperms (monocotyledonous and dicotyledonous plants), gymnosperms, ferns, horsetails, psilophytes, lycophytes, bryophytes, and multicellular algae. (See for example, Figure 1, adapted from Daly et al. 2001 *Plant Physiology* 127:1328-1333; and see also Tudge, C., The Variety of Life, Oxford University Press, New York, 2000, pp. 547-606.)

A "transgenic plant" refers to a plant that contains genetic material not found in a wild type plant of the same species, variety or cultivar. The genetic material may

include a transgene, an insertional mutagenesis event (such as by transposon or T-DNA insertional mutagenesis), an activation tagging sequence, a mutated sequence, a homologous recombination event or a sequence modified by chimeraplasty. Typically, the foreign genetic material has been introduced into the plant by human manipulation, but any method can be used as one of skill in the art recognizes.

A transgenic plant may contain an expression vector or cassette. The expression cassette typically comprises a polypeptide-encoding sequence operably linked (i.e., under regulatory control of) to appropriate inducible or constitutive regulatory sequences that allow for the expression of polypeptide. The expression cassette can be introduced into a plant by transformation or by breeding after transformation of a parent plant. A plant refers to a whole plant as well as to a plant part, such as seed, fruit, leaf, or root, plant tissue, plant cells or any other plant material, e.g., a plant explant, as well as to progeny thereof, and to *in vitro* systems that mimic biochemical or cellular components or processes in a cell.

"Ectopic expression or altered expression" in reference to a polynucleotide indicates that the pattern of expression in, e.g., a transgenic plant or plant tissue, is different from the expression pattern in a wild type plant or a reference plant of the same species. The pattern of expression may also be compared with a reference expression pattern in a wild type plant of the same species. For example, the polynucleotide or polypeptide is expressed in a cell or tissue type other than a cell or tissue type in which the sequence is expressed in the wild type plant, or by expression at a time other than at the time the sequence is expressed in the wild type plant, or by a response to different inducible agents, such as hormones or environmental signals, or at different expression levels (either higher or lower) compared with those found in a wild type plant. The term also refers to altered expression patterns that are produced by lowering the levels of expression to below the detection level or completely abolishing expression. The resulting expression pattern can be transient or stable, constitutive or inducible. In reference to a polypeptide, the term "ectopic expression or altered expression" further may relate to altered activity levels resulting from the interactions of the polypeptides with exogenous or endogenous modulators or from interactions with factors or as a result of the chemical modification of the polypeptides.

A "fragment" or "domain," with respect to a polypeptide, refers to a subsequence of the polypeptide. In some cases, the fragment or domain, is a subsequence of the polypeptide which performs at least one biological function of the intact polypeptide in substantially the same manner, or to a similar extent, as does the intact polypeptide. For example, a polypeptide fragment can comprise a recognizable structural motif or functional domain such as a DNA-binding site or domain that binds to a DNA promoter region, an activation domain, or a domain for protein-protein interactions. Fragments can vary in size from as few as 6 amino acids to the full length of the intact polypeptide, but are preferably at least about 30 amino acids in length and more preferably at least about 60 amino acids in length. In reference to a polynucleotide sequence, "a fragment" refers to any subsequence of a polynucleotide, typically, of at least about 15 consecutive nucleotides, preferably at least about 30 nucleotides, more preferably at least about 50 nucleotides, of any of the sequences provided herein.

The invention also encompasses production of DNA sequences that encode transcription factors and transcription factor derivatives, or fragments thereof, entirely by synthetic chemistry. After production, the synthetic sequence may be inserted into any of the many available expression vectors and cell systems using reagents well known in the art. Moreover, synthetic chemistry may be used to introduce mutations into a sequence encoding transcription factors or any fragment thereof.

A "conserved domain", with respect to a polypeptide, refers to a domain within a transcription factor family which exhibits a higher degree of sequence homology, such as at least 65% sequence identity including conservative substitutions, and preferably at least 80% sequence identity, and more preferably at least 85%, or at least about 86%, or at least about 87%, or at least about 88%, or at least about 90%, or at least about 95%, or at least about 98% amino acid residue sequence identity of a polypeptide of consecutive amino acid residues. A fragment or domain can be referred to as outside a consensus sequence or outside a consensus DNA-binding site that is known to exist or that exists for a particular transcription factor class, family, or sub-family. In this case, the fragment or domain will not include the exact amino acids of a consensus sequence or consensus DNA-binding

site of a transcription factor class, family or sub-family, or the exact amino acids of a particular transcription factor consensus sequence or consensus DNA-binding site. Furthermore, a particular fragment, region, or domain of a polypeptide, or a polynucleotide encoding a polypeptide, can be “outside a conserved domain” if all the amino acids of the fragment, region, or domain fall outside of a defined conserved domain(s) for a polypeptide or protein. The conserved domains for each of polypeptides of SEQ ID NOs:2 - 2N, where N = 2-561, are listed in Table 4 as described in Example VII. Also, many of the polypeptides of Table 4 have conserved domains specifically indicated by start and stop sites. A comparison of the regions of the polypeptides in SEQ ID NOs:2 - 2N, where N = 2-561, or of those in Table 4, allows one of skill in the art to identify conserved domain(s) for any of the polypeptides listed or referred to in this disclosure, including those in Table 4.

A “trait” refers to a physiological, morphological, biochemical, or physical characteristic of a plant or particular plant material or cell. In some instances, this characteristic is visible to the human eye, such as seed or plant size, or can be measured by biochemical techniques, such as detecting the protein, starch, or oil content of seed or leaves, or by observation of a metabolic or physiological process, e.g. by measuring uptake of carbon dioxide, or by the observation of the expression level of a gene or genes, e.g., by employing Northern analysis, RT-PCR, microarray gene expression assays, or reporter gene expression systems, or by agricultural observations such as stress tolerance, yield, or pathogen tolerance. Any technique can be used to measure the amount of, comparative level of, or difference in any selected chemical compound or macromolecule in the transgenic plants, however.

“Trait modification” refers to a detectable difference in a characteristic in a plant ectopically expressing a polynucleotide or polypeptide of the present invention relative to a plant not doing so, such as a wild type plant. In some cases, the trait modification can be evaluated quantitatively. For example, the trait modification can entail at least about a 2% increase or decrease in an observed trait (difference), at least a 5% difference, at least about a 10% difference, at least about a 20% difference, at least about a 30%, at least about a 50%, at least about a 70%, or at least about a 100%, or an even greater difference compared with a wild type plant. It is known that there can be a natural variation in the modified trait. Therefore, the trait modification

observed entails a change of the normal distribution of the trait in the plants compared with the distribution observed in wild type plant.

I. Traits Which May Be Modified

Trait modifications of particular interest include those to seed (such as embryo or endosperm), fruit, root, flower, leaf, stem, shoot, seedling or the like, including: enhanced tolerance to environmental conditions including freezing, chilling, heat, drought, water saturation, radiation and ozone; improved tolerance to microbial, fungal or viral diseases; improved tolerance to pest infestations, including nematodes, mollicutes, parasitic higher plants or the like; decreased herbicide sensitivity; improved tolerance of heavy metals or enhanced ability to take up heavy metals; improved growth under poor photoconditions (e.g., low light and/or short day length), or changes in expression levels of genes of interest. Other phenotype that can be modified relate to the production of plant metabolites, such as variations in the production of taxol, tocopherol, tocotrienol, sterols, phytosterols, vitamins, wax monomers, anti-oxidants, amino acids, lignins, cellulose, tannins, prenyl lipids (such as chlorophylls and carotenoids), glucosinolates, and terpenoids, enhanced or compositionally altered protein or oil production (especially in seeds), or modified sugar (insoluble or soluble) and/or starch composition. Physical plant characteristics that can be modified include cell development (such as the number of trichomes), fruit and seed size and number, yields of plant parts such as stems, leaves, inflorescences, and roots, the stability of the seeds during storage, characteristics of the seed pod (e.g., susceptibility to shattering), root hair length and quantity, internode distances, or the quality of seed coat. Plant growth characteristics that can be modified include growth rate, germination rate of seeds, vigor of plants and seedlings, leaf and flower senescence, male sterility, apomixis, flowering time, flower abscission, rate of nitrogen uptake, osmotic sensitivity to soluble sugar concentrations, biomass or transpiration characteristics, as well as plant architecture characteristics such as apical dominance, branching patterns, number of organs, organ identity, organ shape or size.

II. Transcription Factors Modify Expression Of Endogenous Genes

Expression of genes which encode transcription factors that modify expression of endogenous genes, polynucleotides, and proteins are well known in the art. In addition, transgenic plants comprising isolated polynucleotides encoding transcription

factors may also modify expression of endogenous genes, polynucleotides, and proteins. Examples include Peng et al. (1997, *Genes and Development* 11:3194-3205) and Peng et al. (1999, *Nature*, 400:256-261). In addition, many others have demonstrated that an *Arabidopsis* transcription factor expressed in an exogenous plant species elicits the same or very similar phenotypic response. See, for example, Fu et al. (2001, *Plant Cell* 13:1791-1802); Nandi et al. (2000, *Curr. Biol.* 10:215-218); Coupland (1995, *Nature* 377:482-483); and Weigel and Nilsson (1995, *Nature* 377:482-500).

In another example, Mandel et al. (1992, *Cell* 71:133-143) and Suzuki et al. (2001, *Plant J.* 28:409-418) teach that a transcription factor expressed in another plant species elicits the same or very similar phenotypic response of the endogenous sequence, as often predicted in earlier studies of *Arabidopsis* transcription factors in *Arabidopsis* (see Mandel et al., 1992, *supra*; Suzuki et al., 2001, *supra*).

Other examples include Müller et al. (2001, *Plant J.* 28:169-179); Kim et al. (2001, *Plant J.* 25:247-259); Kyoizuka and Shimamoto (2002, *Plant Cell Physiol.* 43:130-135); Boss and Thomas (2002, *Nature*, 416:847-850); He et al. (2000, *Transgenic Res.*, 9:223-227); and Robson et al. (2001, *Plant J.* 28:619-631).

In yet another example, Gilmour et al. (1998, *Plant J.* 16:433-442) teach an *Arabidopsis* AP2 transcription factor, CBF1, which, when overexpressed in transgenic plants, increases plant freezing tolerance. Jaglo et al (2001, *Plant Physiol.* 127:910-917) further identified sequences in *Brassica napus* which encode CBF-like genes and that transcripts for these genes accumulated rapidly in response to low temperature. Transcripts encoding CBF-like proteins were also found to accumulate rapidly in response to low temperature in wheat, as well as in tomato. An alignment of the CBF proteins from *Arabidopsis*, *B. napus*, wheat, rye, and tomato revealed the presence of conserved amino acid sequences, PKK/RPAGR_xKFxETRHP and DSAWR, that bracket the AP2/EREBP DNA binding domains of the proteins and distinguish them from other members of the AP2/EREBP protein family. (See Jaglo et al., *supra*.)

III. Polypeptides and Polynucleotides of the Invention

The present invention provides, among other things, transcription factors (TFs), and transcription factor homologue polypeptides, and isolated or recombinant polynucleotides encoding the polypeptides, or novel variant polypeptides or polynucleotides encoding novel variants of transcription factors derived from the specific sequences provided here. These polypeptides and polynucleotides may be employed to modify a plant's characteristic.

Exemplary polynucleotides encoding the polypeptides of the invention were identified in the *Arabidopsis thaliana* GenBank database using publicly available sequence analysis programs and parameters. Sequences initially identified were then further characterized to identify sequences comprising specified sequence strings corresponding to sequence motifs present in families of known transcription factors. In addition, further exemplary polynucleotides encoding the polypeptides of the invention were identified in the plant GenBank database using publicly available sequence analysis programs and parameters. Sequences initially identified were then further characterized to identify sequences comprising specified sequence strings corresponding to sequence motifs present in families of known transcription factors. Polynucleotide sequences meeting such criteria were confirmed as transcription factors.

Additional polynucleotides of the invention were identified by screening *Arabidopsis thaliana* and/or other plant cDNA libraries with probes corresponding to known transcription factors under low stringency hybridization conditions. Additional sequences, including full length coding sequences were subsequently recovered by the rapid amplification of cDNA ends (RACE) procedure, using a commercially available kit according to the manufacturer's instructions. Where necessary, multiple rounds of RACE are performed to isolate 5' and 3' ends. The full length cDNA was then recovered by a routine end-to-end polymerase chain reaction (PCR) using primers specific to the isolated 5' and 3' ends. Exemplary sequences are provided in the Sequence Listing.

The polynucleotides of the invention can be or were ectopically expressed in overexpressor or knockout plants and the changes in the characteristic(s) or trait(s) of

the plants observed. Therefore, the polynucleotides and polypeptides can be employed to improve the characteristics of plants.

The polynucleotides of the invention can be or were ectopically expressed in overexpressor plant cells and the changes in the expression levels of a number of genes, polynucleotides, and/or proteins of the plant cells observed. Therefore, the polynucleotides and polypeptides can be employed to change expression levels of a genes, polynucleotides, and/or proteins of plants.

IV. Producing Polypeptides

The polynucleotides of the invention include sequences that encode transcription factors and transcription factor homologue polypeptides and sequences complementary thereto, as well as unique fragments of coding sequence, or sequence complementary thereto. Such polynucleotides can be, e.g., DNA or RNA, e.g., mRNA, cRNA, synthetic RNA, genomic DNA, cDNA synthetic DNA, oligonucleotides, etc. The polynucleotides are either double-stranded or single-stranded, and include either, or both sense (i.e., coding) sequences and antisense (i.e., non-coding, complementary) sequences. The polynucleotides include the coding sequence of a transcription factor, or transcription factor homologue polypeptide, in isolation, in combination with additional coding sequences (e.g., a purification tag, a localization signal, as a fusion-protein, as a pre-protein, or the like), in combination with non-coding sequences (e.g., introns or inteins, regulatory elements such as promoters, enhancers, terminators, and the like), and/or in a vector or host environment in which the polynucleotide encoding a transcription factor or transcription factor homologue polypeptide is an endogenous or exogenous gene.

A variety of methods exist for producing the polynucleotides of the invention. Procedures for identifying and isolating DNA clones are well known to those of skill in the art, and are described in, e.g., Berger and Kimmel, Guide to Molecular Cloning Techniques, Methods in Enzymology volume 152 Academic Press, Inc., San Diego, CA ("Berger"); Sambrook et al., Molecular Cloning - A Laboratory Manual (2nd Ed.), Vol. 1-3, Cold Spring Harbor Laboratory, Cold Spring Harbor, New York, 1989 ("Sambrook") and Current Protocols in Molecular Biology, F. M. Ausubel et al., eds.,

Current Protocols, a joint venture between Greene Publishing Associates, Inc. and John Wiley & Sons, Inc., (supplemented through 2000) ("Ausubel").

Alternatively, polynucleotides of the invention, can be produced by a variety of in vitro amplification methods adapted to the present invention by appropriate selection of specific or degenerate primers. Examples of protocols sufficient to direct persons of skill through in vitro amplification methods, including the polymerase chain reaction (PCR) the ligase chain reaction (LCR), Qbeta-replicase amplification and other RNA polymerase mediated techniques (e.g., NASBA), e.g., for the production of the homologous nucleic acids of the invention are found in Berger (*supra*), Sambrook (*supra*), and Ausubel (*supra*), as well as Mullis et al., (1987) PCR Protocols A Guide to Methods and Applications (Innis et al. eds) Academic Press Inc. San Diego, CA (1990) (Innis). Improved methods for cloning in vitro amplified nucleic acids are described in Wallace et al., U.S. Pat. No. 5,426,039. Improved methods for amplifying large nucleic acids by PCR are summarized in Cheng et al. (1994) Nature 369: 684-685 and the references cited therein, in which PCR amplicons of up to 40kb are generated. One of skill will appreciate that essentially any RNA can be converted into a double stranded DNA suitable for restriction digestion, PCR expansion and sequencing using reverse transcriptase and a polymerase. *See, e.g.,* Ausubel, Sambrook and Berger, *all supra*.

Alternatively, polynucleotides and oligonucleotides of the invention can be assembled from fragments produced by solid-phase synthesis methods. Typically, fragments of up to approximately 100 bases are individually synthesized and then enzymatically or chemically ligated to produce a desired sequence, e.g., a polynucleotide encoding all or part of a transcription factor. For example, chemical synthesis using the phosphoramidite method is described, e.g., by Beaucage et al. (1981) Tetrahedron Letters 22:1859-1869; and Matthes et al. (1984) EMBO J. 3:801-805. According to such methods, oligonucleotides are synthesized, purified, annealed to their complementary strand, ligated and then optionally cloned into suitable vectors. And if so desired, the polynucleotides and polypeptides of the invention can be custom ordered from any of a number of commercial suppliers.

V. Homologous Sequences

Sequences homologous, i.e., that share significant sequence identity or similarity, to those provided in the Sequence Listing, derived from *Arabidopsis thaliana* or from other plants of choice are also an aspect of the invention.

Homologous sequences can be derived from any plant including monocots and dicots and in particular agriculturally important plant species, including but not limited to, crops such as soybean, wheat, corn, potato, cotton, rice, rape, oilseed rape (including canola), sunflower, alfalfa, sugarcane and turf; or fruits and vegetables, such as banana, blackberry, blueberry, strawberry, and raspberry, cantaloupe, carrot, cauliflower, coffee, cucumber, eggplant, grapes, honeydew, lettuce, mango, melon, onion, papaya, peas, peppers, pineapple, pumpkin, spinach, squash, sweet corn, tobacco, tomato, watermelon, rosaceous fruits (such as apple, peach, pear, cherry and plum) and vegetable brassicas (such as broccoli, cabbage, cauliflower, Brussels sprouts, and kohlrabi). Other crops, fruits and vegetables whose phenotype can be changed include barley, rye, millet, sorghum, currant, avocado, citrus fruits such as oranges, lemons, grapefruit and tangerines, artichoke, cherries, nuts such as the walnut and peanut, endive, leek, roots, such as arrowroot, beet, cassava, turnip, radish, yam, and sweet potato, and beans. The homologous sequences may also be derived from woody species, such as pine, poplar and eucalyptus, or mint or other labiates.

Orthologs And Paralogs

Several different methods are known by those of skill in the art for identifying and defining these functionally homologous sequences. Three general methods for defining paralogs and orthologs are described; a paralog or ortholog or homolog may be identified by one or more of the methods described below.

Orthologs and paralogs are evolutionarily related genes that have similar sequence and similar functions. Orthologs are structurally related genes in different species that are derived from a speciation event. Paralogs are structurally related genes within a single species that are derived by a duplication event.

Within a single plant species, gene duplication may cause two copies of a particular gene, giving rise to two or more genes with similar sequence and similar function known as paralogs. A paralog is therefore a similar gene with a similar function within the same species. Paralogs typically cluster together or in the same

clade (a group of similar genes) when a gene family phylogeny is analyzed using programs such as CLUSTAL (Thompson et al. (1994) *Nucleic Acids Res.* 22:4673-4680; Higgins et al. (1996) *Methods Enzymol.* 266 383-402). Groups of similar genes can also be identified with pair-wise BLAST analysis (Feng and Doolittle (1987) *J. Mol. Evol.* 25:351-360). For example, a clade of very similar MADS domain transcription factors from *Arabidopsis* all share a common function in flowering time (Ratcliffe et al. (2001) *Plant Physiol.* 126:122-132), and a group of very similar AP2 domain transcription factors from *Arabidopsis* are involved in tolerance of plants to freezing (Gilmour et al. (1998) *Plant J.* 16:433-442). Analysis of groups of similar genes with similar function that fall within one clade can yield sub-sequences that are particular to the clade. These sub-sequences, known as consensus sequences, can not only be used to define the sequences within each clade, but define the functions of these genes; genes within a clade may contain paralogous or orthologous sequences that share the same function. (See also, for example, Mount, D.W. (2001) Bioinformatics: Sequence and Genome Analysis Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York page 543.)

Speciation, the production of new species from a parental species, can also give rise to two or more genes with similar sequence and similar function. These genes, termed orthologs, often have an identical function within their host plants and are often interchangeable between species without losing function. Because plants have common ancestors, many genes in any plant species will have a corresponding orthologous gene in another plant species. Once a phylogenetic tree for a gene family of one species has been constructed using a program such as CLUSTAL (Thompson et al. (1994) *Nucleic Acids Res.* 22:4673-4680; Higgins et al. (1996) *Methods Enzymol.* 266:383-402), potential orthologous sequences can be placed into the phylogenetic tree and its relationship to genes from the species of interest can be determined. Once the ortholog pair has been identified, the function of the test ortholog can be determined by determining the function of the reference ortholog.

Transcription factors that are homologous to the listed sequences will typically share at least about 30% amino acid sequence identity, or at least about 30% amino acid sequence identity outside of a known consensus sequence or consensus DNA-binding site. More closely related transcription factors can share at least about 50%,

about 60%, about 65%, about 70%, about 75% or about 80% or about 90% or about 95% or about 98% or more sequence identity with the listed sequences, or with the listed sequences but excluding or outside a known consensus sequence or consensus DNA-binding site, or with the listed sequences excluding one or all conserved domain. Factors that are most closely related to the listed sequences share, e.g., at least about 85%, about 90% or about 95% or more % sequence identity to the listed sequences, or to the listed sequences but excluding or outside a known consensus sequence or consensus DNA-binding site or outside one or all conserved domain. At the nucleotide level, the sequences will typically share at least about 40% nucleotide sequence identity, preferably at least about 50%, about 60%, about 70% or about 80% sequence identity, and more preferably about 85%, about 90%, about 95% or about 97% or more sequence identity to one or more of the listed sequences, or to a listed sequence but excluding or outside a known consensus sequence or consensus DNA-binding site, or outside one or all conserved domain. The degeneracy of the genetic code enables major variations in the nucleotide sequence of a polynucleotide while maintaining the amino acid sequence of the encoded protein. Conserved domains within a transcription factor family may exhibit a higher degree of sequence homology, such as at least 65% sequence identity including conservative substitutions, and preferably at least 80% sequence identity, and more preferably at least 85%, or at least about 86%, or at least about 87%, or at least about 88%, or at least about 90%, or at least about 95%, or at least about 98% sequence identity. Transcription factors that are homologous to the listed sequences should share at least 30%, or at least about 60%, or at least about 75%, or at least about 80%, or at least about 90%, or at least about 95% amino acid sequence identity over the entire length of the polypeptide or the homolog. In addition, transcription factors that are homologous to the listed sequences should share at least 30%, or at least about 60%, or at least about 75%, or at least about 80%, or at least about 90%, or at least about 95% amino acid sequence similarity over the entire length of the polypeptide or the homolog.

Percent identity can be determined electronically, e.g., by using the MEGALIGN program (DNASTAR, Inc. Madison, Wis.). The MEGALIGN program can create alignments between two or more sequences according to different methods, e.g., the clustal method. (See, e.g., Higgins, D. G. and P. M. Sharp (1988) *Gene*

73:237-244.) The clustal algorithm groups sequences into clusters by examining the distances between all pairs. The clusters are aligned pairwise and then in groups. Other alignment algorithms or programs may be used, including FASTA, BLAST, or ENTREZ, FASTA and BLAST. These are available as a part of the GCG sequence analysis package (University of Wisconsin, Madison, Wis.), and can be used with or without default settings. ENTREZ is available through the National Center for Biotechnology Information. In one embodiment, the percent identity of two sequences can be determined by the GCG program with a gap weight of 1, e.g., each amino acid gap is weighted as if it were a single amino acid or nucleotide mismatch between the two sequences (see USPN 6,262,333).

Other techniques for alignment are described in *Methods in Enzymology*, vol. 266: *Computer Methods for Macromolecular Sequence Analysis* (1996), ed. Doolittle, Academic Press, Inc., San Diego, Calif., USA. Preferably, an alignment program that permits gaps in the sequence is utilized to align the sequences. The Smith-Waterman is one type of algorithm that permits gaps in sequence alignments. See *Methods Mol. Biol.* 70: 173-187 (1997). Also, the GAP program using the Needleman and Wunsch alignment method can be utilized to align sequences. An alternative search strategy uses MPSRCH software, which runs on a MASPAR computer. MPSRCH uses a Smith-Waterman algorithm to score sequences on a massively parallel computer. This approach improves ability to pick up distantly related matches, and is especially tolerant of small gaps and nucleotide sequence errors. Nucleic acid-encoded amino acid sequences can be used to search both protein and DNA databases.

The percentage similarity between two polypeptide sequences, e.g., sequence A and sequence B, is calculated by dividing the length of sequence A, minus the number of gap residues in sequence A, minus the number of gap residues in sequence B, into the sum of the residue matches between sequence A and sequence B, times one hundred. Gaps of low or of no similarity between the two amino acid sequences are not included in determining percentage similarity. Percent identity between polynucleotide sequences can also be counted or calculated by other methods known in the art, e.g., the Jotun Hein method. (See, e.g., Hein, J. (1990) *Methods Enzymol.* 183:626-645.) Identity between sequences can also be determined by other methods

known in the art, e.g., by varying hybridization conditions (see US Patent Application No. 20010010913).

Thus, the invention provides methods for identifying a sequence similar or paralogous or orthologous or homologous to one or more polynucleotides as noted herein, or one or more target polypeptides encoded by the polynucleotides, or otherwise noted herein and may include linking or associating a given plant phenotype or gene function with a sequence. In the methods, a sequence database is provided (locally or across an inter or intra net) and a query is made against the sequence database using the relevant sequences herein and associated plant phenotypes or gene functions.

In addition, one or more polynucleotide sequences or one or more polypeptides encoded by the polynucleotide sequences may be used to search against a BLOCKS (Bairoch et al. (1997) *Nucleic Acids Res.* 25:217-221), PFAM, and other databases which contain previously identified and annotated motifs, sequences and gene functions. Methods that search for primary sequence patterns with secondary structure gap penalties (Smith et al. (1992) *Protein Engineering* 5:35-51) as well as algorithms such as Basic Local Alignment Search Tool (BLAST; Altschul, S. F. (1993) *J. Mol. Evol.* 36:290-300; Altschul et al. (1990) *supra*), BLOCKS (Henikoff, S. and Henikoff, G. J. (1991) *Nucleic Acids Research* 19:6565-6572), Hidden Markov Models (HMM; Eddy, S. R. (1996) *Cur. Opin. Str. Biol.* 6:361-365; Sonnhammer et al. (1997) *Proteins* 28:405-420), and the like, can be used to manipulate and analyze polynucleotide and polypeptide sequences encoded by polynucleotides. These databases, algorithms and other methods are well known in the art and are described in Ausubel et al. (1997; *Short Protocols in Molecular Biology*, John Wiley & Sons, New York N.Y., unit 7.7) and in Meyers, R. A. (1995; *Molecular Biology and Biotechnology*, Wiley VCH, New York N.Y., p 856-853).

Furthermore, methods using manual alignment of sequences similar or homologous to one or more polynucleotide sequences or one or more polypeptides encoded by the polynucleotide sequences may be used to identify regions of similarity and conserved domains. Such manual methods are well-known of those of skill in the art and can include, for example, comparisons of tertiary structure between a

polypeptide sequence encoded by a polynucleotide which comprises a known function with a polypeptide sequence encoded by a polynucleotide sequence which has a function not yet determined. Such examples of tertiary structure may comprise predicted alpha helices, beta-sheets, amphipathic helices, leucine zipper motifs, zinc finger motifs, proline-rich regions, cysteine repeat motifs, and the like.

VI. Identifying Polynucleotides or Nucleic Acids by Hybridization

Polynucleotides homologous to the sequences illustrated in the Sequence Listing and tables can be identified, e.g., by hybridization to each other under stringent or under highly stringent conditions. Single stranded polynucleotides hybridize when they associate based on a variety of well characterized physical-chemical forces, such as hydrogen bonding, solvent exclusion, base stacking and the like. The stringency of a hybridization reflects the degree of sequence identity of the nucleic acids involved, such that the higher the stringency, the more similar are the two polynucleotide strands. Stringency is influenced by a variety of factors, including temperature, salt concentration and composition, organic and non-organic additives, solvents, etc. present in both the hybridization and wash solutions and incubations (and number thereof), as described in more detail in the references cited above. Encompassed by the invention are polynucleotide sequences that are capable of hybridizing to the claimed polynucleotide sequences, and, in particular, to those shown in SEQ ID NOs: 860; 802; 240; 274; 558; 24; 1120; 44; 460; 286; 120; 130; 134; 698; 832; 580; 612; 48, and fragments thereof under various conditions of stringency. (See, e.g., Wahl, G. M. and S. L. Berger (1987) *Methods Enzymol.* 152:399-407; Kimmel, A. R. (1987) *Methods Enzymol.* 152:507-511.) Estimates of homology are provided by either DNA-DNA or DNA-RNA hybridization under conditions of stringency as is well understood by those skilled in the art (Hames and Higgins, Eds. (1985) *Nucleic Acid Hybridisation*, IRL Press, Oxford, U.K.). Stringency conditions can be adjusted to screen for moderately similar fragments, such as homologous sequences from distantly related organisms, to highly similar fragments, such as genes that duplicate functional enzymes from closely related organisms. Post-hybridization washes determine stringency conditions.

In addition to the nucleotide sequences listed in Tables 4 and 5, full length cDNA, orthologs, paralogs and homologs of the present nucleotide sequences may be

identified and isolated using well known methods. The cDNA libraries orthologs, paralogs and homologs of the present nucleotide sequences may be screened using hybridization methods to determine their utility as hybridization target or amplification probes.

An example of stringent hybridization conditions for hybridization of complementary nucleic acids which have more than 100 complementary residues on a filter in a Southern or northern blot is about 5°C to 20°C lower than the thermal melting point (T_m) for the specific sequence at a defined ionic strength and pH. The T_m is the temperature (under defined ionic strength and pH) at which 50% of the target sequence hybridizes to a perfectly matched probe. Nucleic acid molecules that hybridize under stringent conditions will typically hybridize to a probe based on either the entire cDNA or selected portions, e.g., to a unique subsequence, of the cDNA under wash conditions of 0.2x SSC to 2.0 x SSC, 0.1% SDS at 50-65° C. For example, high stringency is about 0.2 x SSC, 0.1% SDS at 65° C. Ultra-high stringency will be the same conditions except the wash temperature is raised about 3 to about 5° C, and ultra-ultra-high stringency will be the same conditions except the wash temperature is raised about 6 to about 9° C. For identification of less closely related homologues washes can be performed at a lower temperature, e.g., 50° C. In general, stringency is increased by raising the wash temperature and/or decreasing the concentration of SSC, as known in the art.

In another example, stringent salt concentration will ordinarily be less than about 750 mM NaCl and 75 mM trisodium citrate, preferably less than about 500 mM NaCl and 50 mM trisodium citrate, and most preferably less than about 250 mM NaCl and 25 mM trisodium citrate. Low stringency hybridization can be obtained in the absence of organic solvent, e.g., formamide, while high stringency hybridization can be obtained in the presence of at least about 35% formamide, and most preferably at least about 50% formamide. Stringent temperature conditions will ordinarily include temperatures of at least about 30° C, more preferably of at least about 37° C, and most preferably of at least about 42° C. Varying additional parameters, such as hybridization time, the concentration of detergent, e.g., sodium dodecyl sulfate (SDS), and the inclusion or exclusion of carrier DNA, are well known to those skilled in the

art. Various levels of stringency are accomplished by combining these various conditions as needed. In a preferred embodiment, hybridization will occur at 30° C in 750 mM NaCl, 75 mM trisodium citrate, and 1% SDS. In a more preferred embodiment, hybridization will occur at 37° C in 500 mM NaCl, 50 mM trisodium citrate, 1% SDS, 35% formamide, and 100 µg/ml denatured salmon sperm DNA (ssDNA). In a most preferred embodiment, hybridization will occur at 42° C in 250 mM NaCl, 25 mM trisodium citrate, 1% SDS, 50% formamide, and 200 µg/ml ssDNA. Useful variations on these conditions will be readily apparent to those skilled in the art.

The washing steps that follow hybridization can also vary in stringency. Wash stringency conditions can be defined by salt concentration and by temperature. As above, wash stringency can be increased by decreasing salt concentration or by increasing temperature. For example, stringent salt concentration for the wash steps will preferably be less than about 30 mM NaCl and 3 mM trisodium citrate, and most preferably less than about 15 mM NaCl and 1.5 mM trisodium citrate. Stringent temperature conditions for the wash steps will ordinarily include temperature of at least about 25° C, more preferably of at least about 42° C. Another preferred set of highly stringent conditions uses two final washes in 0.1X SSC, 0.1% SDS at 65° C. The most preferred high stringency washes are of at least about 68° C. For example, in a preferred embodiment, wash steps will occur at 25° C in 30 mM NaCl, 3 mM trisodium citrate, and 0.1% SDS. In a more preferred embodiment, wash steps will occur at 42° C in 15 mM NaCl, 1.5 mM trisodium citrate, and 0.1% SDS. In a most preferred embodiment, the wash steps will occur at 68° C in 15 mM NaCl, 1.5 mM trisodium citrate, and 0.1% SDS. Additional variations on these conditions will be readily apparent to those skilled in the art (see U.S. Patent Application No. 20010010913).

As another example, stringent conditions can be selected such that an oligonucleotide that is perfectly complementary to the coding oligonucleotide hybridizes to the coding oligonucleotide with at least about a 5-10x higher signal to noise ratio than the ratio for hybridization of the perfectly complementary oligonucleotide to a nucleic acid encoding a transcription factor known as of the filing

date of the application. Conditions can be selected such that a higher signal to noise ratio is observed in the particular assay which is used, e.g., about 15x, 25x, 35x, 50x or more. Accordingly, the subject nucleic acid hybridizes to the unique coding oligonucleotide with at least a 2x higher signal to noise ratio as compared to hybridization of the coding oligonucleotide to a nucleic acid encoding known polypeptide. Again, higher signal to noise ratios can be selected, e.g., about 5x, 10x, 25x, 35x, 50x or more. The particular signal will depend on the label used in the relevant assay, e.g., a fluorescent label, a colorimetric label, a radioactive label, or the like.

Alternatively, transcription factor homolog polypeptides can be obtained by screening an expression library using antibodies specific for one or more transcription factors. With the provision herein of the disclosed transcription factor, and transcription factor homologue nucleic acid sequences, the encoded polypeptide(s) can be expressed and purified in a heterologous expression system (e.g., *E. coli*) and used to raise antibodies (monoclonal or polyclonal) specific for the polypeptide(s) in question. Antibodies can also be raised against synthetic peptides derived from transcription factor, or transcription factor homologue, amino acid sequences. Methods of raising antibodies are well known in the art and are described in Harlow and Lane (1988) Antibodies: A Laboratory Manual, Cold Spring Harbor Laboratory, New York. Such antibodies can then be used to screen an expression library produced from the plant from which it is desired to clone additional transcription factor homologues, using the methods described above. The selected cDNAs can be confirmed by sequencing and enzymatic activity.

VII. Sequence Variations

It will readily be appreciated by those of skill in the art, that any of a variety of polynucleotide sequences are capable of encoding the transcription factors and transcription factor homologue polypeptides of the invention. Due to the degeneracy of the genetic code, many different polynucleotides can encode identical and/or substantially similar polypeptides in addition to those sequences illustrated in the Sequence Listing. Nucleic acids having a sequence that differs from the sequences shown in the Sequence Listing, or complementary sequences, that encode functionally equivalent peptides (i.e., peptides having some degree of equivalent or similar

biological activity) but differ in sequence from the sequence shown in the sequence listing due to degeneracy in the genetic code, are also within the scope of the invention.

Altered polynucleotide sequences encoding polypeptides include those sequences with deletions, insertions, or substitutions of different nucleotides, resulting in a polynucleotide encoding a polypeptide with at least one functional characteristic of the instant polypeptides. Included within this definition are polymorphisms which may or may not be readily detectable using a particular oligonucleotide probe of the polynucleotide encoding the instant polypeptides, and improper or unexpected hybridization to allelic variants, with a locus other than the normal chromosomal locus for the polynucleotide sequence encoding the instant polypeptides.

Allelic variant refers to any of two or more alternative forms of a gene occupying the same chromosomal locus. Allelic variation arises naturally through mutation, and may result in phenotypic polymorphism within populations. Gene mutations can be silent (i.e., no change in the encoded polypeptide) or may encode polypeptides having altered amino acid sequence. The term allelic variant is also used herein to denote a protein encoded by an allelic variant of a gene. Splice variant refers to alternative forms of RNA transcribed from a gene. Splice variation arises naturally through use of alternative splicing sites within a transcribed RNA molecule, or less commonly between separately transcribed RNA molecules, and may result in several mRNAs transcribed from the same gene. Splice variants may encode polypeptides having altered amino acid sequence. The term splice variant is also used herein to denote a protein encoded by a splice variant of an mRNA transcribed from a gene.

Those skilled in the art would recognize that the polypeptide sequence G681, SEQ ID NO: 580, represents a single transcription factor; allelic variation and alternative splicing may be expected to occur. Allelic variants of the polypeptide sequence of SEQ ID NO: 579 can be cloned by probing cDNA or genomic libraries from different individual organisms according to standard procedures. Allelic variants of the DNA sequence shown in SEQ ID NO: 579, including those containing silent mutations and those in which mutations result in amino acid sequence changes, are within the scope of the present invention, as are proteins which are allelic variants

of SEQ ID NO: 580. cDNAs generated from alternatively spliced mRNAs, which retain the properties of the transcription factor are included within the scope of the present invention, as are polypeptides encoded by such cDNAs and mRNAs. Allelic variants and splice variants of these sequences can be cloned by probing cDNA or genomic libraries from different individual organisms or tissues according to standard procedures known in the art (see USPN 6,388,064).

For example, Table 1 illustrates, e.g., that the codons AGC, AGT, TCA, TCC, TCG, and TCT all encode the same amino acid: serine. Accordingly, at each position in the sequence where there is a codon encoding serine, any of the above trinucleotide sequences can be used without altering the encoded polypeptide.

Table 1

Amino acid			Possible Codons					
Alanine	Ala	A	GCA	GCC	GCG	GCU		
Cysteine	Cys	C	TGC	TGT				
Aspartic acid	Asp	D	GAC	GAT				
Glutamic acid	Glu	E	GAA	GAG				
Phenylalanine	Phe	F	TTC	TTT				
Glycine	Gly	G	GGA	GGC	GGG	GGT		
Histidine	His	H	CAC	CAT				
Isoleucine	Ile	I	ATA	ATC	ATT			
Lysine	Lys	K	AAA	AAG				
Leucine	Leu	L	TTA	TTG	CTA	CTC	CTG	CTT
Methionine	Met	M	ATG					
Asparagine	Asn	N	AAC	AAT				
Proline	Pro	P	CCA	CCC	CCG	CCT		
Glutamine	Gln	Q	CAA	CAG				
Arginine	Arg	R	AGA	AGG	CGA	CGC	CGG	CGT
Serine	Ser	S	AGC	AGT	TCA	TCC	TCG	TCT
Threonine	Thr	T	ACA	ACC	ACG	ACT		
Valine	Val	V	GTA	GTC	GTG	GTT		
Tryptophan	Trp	W	TGG					
Tyrosine	Tyr	Y	TAC	TAT				

Sequence alterations that do not change the amino acid sequence encoded by the polynucleotide are termed “silent” variations. With the exception of the codons ATG and TGG, encoding methionine and tryptophan, respectively, any of the possible codons for the same amino acid can be substituted by a variety of techniques, e.g., site-directed mutagenesis, available in the art. Accordingly, any and all such variations of a sequence selected from the above table are a feature of the invention.

In addition to silent variations, other conservative variations that alter one, or a few amino acids in the encoded polypeptide, can be made without altering the function of the polypeptide, these conservative variants are, likewise, a feature of the invention.

For example, substitutions, deletions and insertions introduced into the sequences provided in the Sequence Listing are also envisioned by the invention. Such sequence modifications can be engineered into a sequence by site-directed mutagenesis (Wu (ed.) Meth. Enzymol. (1993) vol. 217, Academic Press) or the other methods noted below. Amino acid substitutions are typically of single residues; insertions usually will be on the order of about from 1 to 10 amino acid residues; and deletions will range about from 1 to 30 residues. In preferred embodiments, deletions or insertions are made in adjacent pairs, e.g., a deletion of two residues or insertion of two residues. Substitutions, deletions, insertions or any combination thereof can be combined to arrive at a sequence. The mutations that are made in the polynucleotide encoding the transcription factor should not place the sequence out of reading frame and should not create complementary regions that could produce secondary mRNA structure. Preferably, the polypeptide encoded by the DNA performs the desired function.

Conservative substitutions are those in which at least one residue in the amino acid sequence has been removed and a different residue inserted in its place. Such substitutions generally are made in accordance with the Table 2 when it is desired to maintain the activity of the protein. Table 2 shows amino acids which can be substituted for an amino acid in a protein and which are typically regarded as conservative substitutions.

Table 2

Residue	Conservative Substitutions
Ala	Ser
Arg	Lys
Asn	Gln; His
Asp	Glu
Gln	Asn
Cys	Ser
Glu	Asp
Gly	Pro
His	Asn; Gln
Ile	Leu, Val
Leu	Ile; Val
Lys	Arg; Gln
Met	Leu; Ile
Phe	Met; Leu; Tyr
Ser	Thr; Gly
Thr	Ser; Val
Trp	Tyr
Tyr	Trp; Phe
Val	Ile; Leu

Similar substitutions are those in which at least one residue in the amino acid sequence has been removed and a different residue inserted in its place. Such substitutions generally are made in accordance with the Table 3 when it is desired to maintain the activity of the protein. Table 3 shows amino acids which can be substituted for an amino acid in a protein and which are typically regarded as structural and functional substitutions. For example, a residue in column 1 of Table 3

may be substituted with residue in column 2; in addition, a residue in column 2 of Table 3 may be substituted with the residue of column 1.

Table 3

Residue	Similar Substitutions
Ala	Ser; Thr; Gly; Val; Leu; Ile
Arg	Lys; His; Gly
Asn	Gln; His; Gly; Ser; Thr
Asp	Glu; Ser; Thr
Gln	Asn; Ala
Cys	Ser; Gly
Glu	Asp
Gly	Pro; Arg
His	Asn; Gln; Tyr; Phe; Lys; Arg
Ile	Ala; Leu; Val; Gly; Met
Leu	Ala; Ile; Val; Gly; Met
Lys	Arg; His; Gln; Gly; Pro
Met	Leu; Ile; Phe
Phe	Met; Leu; Tyr; Trp; His; Val; Ala
Ser	Thr; Gly; Asp; Ala; Val; Ile; His
Thr	Ser; Val; Ala; Gly
Trp	Tyr; Phe; His
Tyr	Trp; Phe; His
Val	Ala; Ile; Leu; Gly; Thr; Ser; Glu

Substitutions that are less conservative than those in Table 2 can be selected by picking residues that differ more significantly in their effect on maintaining (a) the structure of the polypeptide backbone in the area of the substitution, for example, as a sheet or helical conformation, (b) the charge or hydrophobicity of the molecule at the target site, or (c) the bulk of the side chain. The substitutions which in general are

expected to produce the greatest changes in protein properties will be those in which (a) a hydrophilic residue, e.g., seryl or threonyl, is substituted for (or by) a hydrophobic residue, e.g., leucyl, isoleucyl, phenylalanyl, valyl or alanyl; (b) a cysteine or proline is substituted for (or by) any other residue; (c) a residue having an electropositive side chain, e.g., lysyl, arginyl, or histidyl, is substituted for (or by) an electronegative residue, e.g., glutamyl or aspartyl; or (d) a residue having a bulky side chain, e.g., phenylalanine, is substituted for (or by) one not having a side chain, e.g., glycine.

VIII. Further Modifying Sequences of the Invention – Mutation/Forced Evolution

In addition to generating silent or conservative substitutions as noted, above, the present invention optionally includes methods of modifying the sequences of the Sequence Listing. In the methods, nucleic acid or protein modification methods are used to alter the given sequences to produce new sequences and/or to chemically or enzymatically modify given sequences to change the properties of the nucleic acids or proteins.

Thus, in one embodiment, given nucleic acid sequences are modified, e.g., according to standard mutagenesis or artificial evolution methods to produce modified sequences. The modified sequences may be created using purified natural polynucleotides isolated from any organism or may be synthesized from purified compositions and chemicals using chemical means well known to those of skill in the art. For example, Ausubel, *supra*, provides additional details on mutagenesis methods. Artificial forced evolution methods are described, for example, by Stemmer (1994) Nature 370:389-391, Stemmer (1994) Proc. Natl. Acad. Sci. USA 91:10747-10751, and U.S. Patents 5,811,238, 5,837,500, and 6,242,568. Methods for engineering synthetic transcription factors and other polypeptides are described, for example, by Zhang et al. (2000) J. Biol. Chem. 275:33850-33860, Liu et al. (2001) J. Biol. Chem. 276:11323-11334, and Isalan et al. (2001) Nature Biotechnol. 19:656-660. Many other mutation and evolution methods are also available and expected to be within the skill of the practitioner.

Similarly, chemical or enzymatic alteration of expressed nucleic acids and polypeptides can be performed by standard methods. For example, sequence can be modified by addition of lipids, sugars, peptides, organic or inorganic compounds, by the inclusion of modified nucleotides or amino acids, or the like. For example, protein modification techniques are illustrated in Ausubel, *supra*. Further details on chemical and enzymatic modifications can be found herein. These modification methods can be used to modify any given sequence, or to modify any sequence produced by the various mutation and artificial evolution modification methods noted herein.

Accordingly, the invention provides for modification of any given nucleic acid by mutation, evolution, chemical or enzymatic modification, or other available methods, as well as for the products produced by practicing such methods, e.g., using the sequences herein as a starting substrate for the various modification approaches.

For example, optimized coding sequence containing codons preferred by a particular prokaryotic or eukaryotic host can be used e.g., to increase the rate of translation or to produce recombinant RNA transcripts having desirable properties, such as a longer half-life, as compared with transcripts produced using a non-optimized sequence. Translation stop codons can also be modified to reflect host preference. For example, preferred stop codons for *Saccharomyces cerevisiae* and mammals are TAA and TGA, respectively. The preferred stop codon for monocotyledonous plants is TGA, whereas insects and *E. coli* prefer to use TAA as the stop codon.

The polynucleotide sequences of the present invention can also be engineered in order to alter a coding sequence for a variety of reasons, including but not limited to, alterations which modify the sequence to facilitate cloning, processing and/or expression of the gene product. For example, alterations are optionally introduced using techniques which are well known in the art, e.g., site-directed mutagenesis, to insert new restriction sites, to alter glycosylation patterns, to change codon preference, to introduce splice sites, etc.

Furthermore, a fragment or domain derived from any of the polypeptides of the invention can be combined with domains derived from other transcription factors or synthetic domains to modify the biological activity of a transcription factor. For instance, a DNA-binding domain derived from a transcription factor of the invention can be combined with the activation domain of another transcription factor or with a synthetic activation domain. A transcription activation domain assists in initiating transcription from a DNA-binding site. Examples include the transcription activation region of VP16 or GAL4 (Moore et al. (1998) Proc. Natl. Acad. Sci. USA 95: 376-381; and Aoyama et al. (1995) Plant Cell 7:1773-1785), peptides derived from bacterial sequences (Ma and Ptashne (1987) Cell 51; 113-119) and synthetic peptides (Giniger and Ptashne, (1987) Nature 330:670-672).

IX. Expression and Modification of Polypeptides

Typically, polynucleotide sequences of the invention are incorporated into recombinant DNA (or RNA) molecules that direct expression of polypeptides of the invention in appropriate host cells, transgenic plants, in vitro translation systems, or the like. Due to the inherent degeneracy of the genetic code, nucleic acid sequences which encode substantially the same or a functionally equivalent amino acid sequence can be substituted for any listed sequence to provide for cloning and expressing the relevant homologue.

X. Vectors, Promoters, and Expression Systems

The present invention includes recombinant constructs comprising one or more of the nucleic acid sequences herein. The constructs typically comprise a vector, such as a plasmid, a cosmid, a phage, a virus (e.g., a plant virus), a bacterial artificial chromosome (BAC), a yeast artificial chromosome (YAC), or the like, into which a nucleic acid sequence of the invention has been inserted, in a forward or reverse orientation. In a preferred aspect of this embodiment, the construct further comprises regulatory sequences, including, for example, a promoter, operably linked to the sequence. Large numbers of suitable vectors and promoters are known to those of skill in the art, and are commercially available.

General texts that describe molecular biological techniques useful herein, including the use and production of vectors, promoters and many other relevant

topics, include Berger, Sambrook and Ausubel, *supra*. Any of the identified sequences can be incorporated into a cassette or vector, e.g., for expression in plants. A number of expression vectors suitable for stable transformation of plant cells or for the establishment of transgenic plants have been described including those described in Weissbach and Weissbach, (1989) Methods for Plant Molecular Biology, Academic Press, and Gelvin et al., (1990) Plant Molecular Biology Manual, Kluwer Academic Publishers. Specific examples include those derived from a Ti plasmid of *Agrobacterium tumefaciens*, as well as those disclosed by Herrera-Estrella et al. (1983) Nature 303: 209, Bevan (1984) Nucl Acid Res. 12: 8711-8721, Klee (1985) Bio/Technology 3: 637-642, for dicotyledonous plants.

Alternatively, non-Ti vectors can be used to transfer the DNA into monocotyledonous plants and cells by using free DNA delivery techniques. Such methods can involve, for example, the use of liposomes, electroporation, microprojectile bombardment, silicon carbide whiskers, and viruses. By using these methods transgenic plants such as wheat, rice (Christou (1991) Bio/Technology 9: 957-962) and corn (Gordon-Kamm (1990) Plant Cell 2: 603-618) can be produced. An immature embryo can also be a good target tissue for monocots for direct DNA delivery techniques by using the particle gun (Weeks et al. (1993) Plant Physiol 102: 1077-1084; Vasil (1993) Bio/Technology 10: 667-674; Wan and Lemeaux (1994) Plant Physiol 104: 37-48, and for *Agrobacterium*-mediated DNA transfer (Ishida et al. (1996) Nature Biotech 14: 745-750).

Typically, plant transformation vectors include one or more cloned plant coding sequence (genomic or cDNA) under the transcriptional control of 5' and 3' regulatory sequences and a dominant selectable marker. Such plant transformation vectors typically also contain a promoter (e.g., a regulatory region controlling inducible or constitutive, environmentally-or developmentally-regulated, or cell- or tissue-specific expression), a transcription initiation start site, an RNA processing signal (such as intron splice sites), a transcription termination site, and/or a polyadenylation signal.

Examples of constitutive plant promoters which can be useful for expressing the TF sequence include: the cauliflower mosaic virus (CaMV) 35S promoter, which

confers constitutive, high-level expression in most plant tissues (*see, e.g.,* Odell et al. (1985) Nature 313:810-812); the nopaline synthase promoter (An et al. (1988) Plant Physiol 88:547-552); and the octopine synthase promoter (Fromm et al. (1989) Plant Cell 1: 977-984).

A variety of plant gene promoters that regulate gene expression in response to environmental, hormonal, chemical, developmental signals, and in a tissue-active manner can be used for expression of a TF sequence in plants. Choice of a promoter is based largely on the phenotype of interest and is determined by such factors as tissue (e.g., seed, fruit, root, pollen, vascular tissue, flower, carpel, etc.), inducibility (e.g., in response to wounding, heat, cold, drought, light, pathogens, etc.), timing, developmental stage, and the like. Numerous known promoters have been characterized and can favorably be employed to promote expression of a polynucleotide of the invention in a transgenic plant or cell of interest. For example, tissue specific promoters include: seed-specific promoters (such as the napin, phaseolin or DC3 promoter described in US Pat. No. 5,773,697), fruit-specific promoters that are active during fruit ripening (such as the *dru1* promoter (US Pat. No. 5,783,393), or the 2A11 promoter (US Pat. No. 4,943,674) and the tomato polygalacturonase promoter (Bird et al. (1988) Plant Mol Biol 11:651), root-specific promoters, such as those disclosed in US Patent Nos. 5,618,988, 5,837,848 and 5,905,186, pollen-active promoters such as PTA29, PTA26 and PTA13 (US Pat. No. 5,792,929), promoters active in vascular tissue (Ringli and Keller (1998) Plant Mol Biol 37:977-988), flower-specific (Kaiser et al. (1995) Plant Mol Biol 28:231-243), pollen (Baerson et al. (1994) Plant Mol Biol 26:1947-1959), carpels (Ohl et al. (1990) Plant Cell 2:837-848), pollen and ovules (Baerson et al. (1993) Plant Mol Biol 22:255-267), auxin-inducible promoters (such as that described in van der Kop et al. (1999) Plant Mol Biol 39:979-990 or Baumann et al. (1999) Plant Cell 11:323-334), cytokinin-inducible promoter (Guevara-Garcia (1998) Plant Mol Biol 38:743-753), promoters responsive to gibberellin (Shi et al. (1998) Plant Mol Biol 38:1053-1060, Willmott et al. (1998) 38:817-825) and the like. Additional promoters are those that elicit expression in response to heat (Ainley et al. (1993) Plant Mol Biol 22: 13-23), light (e.g., the pea *rbcS-3A* promoter, Kuhlemeier et al. (1989) Plant Cell 1:471, and the maize *rbcS* promoter, Schaffner and Sheen (1991) Plant Cell 3: 997); wounding (e.g., *wun1*, Siebertz et al. (1989) Plant Cell 1: 961); pathogens (such as the PR-1

promoter described in Buchel et al. (1999) Plant Mol. Biol. 40:387-396, and the PDF1.2 promoter described in Manners et al. (1998) Plant Mol. Biol. 38:1071-80), and chemicals such as methyl jasmonate or salicylic acid (Gatz et al. (1997) Plant Mol Biol 48: 89-108). In addition, the timing of the expression can be controlled by using promoters such as those acting at senescence (An and Amazon (1995) Science 270: 1986-1988); or late seed development (Odell et al. (1994) Plant Physiol 106:447-458).

Plant expression vectors can also include RNA processing signals that can be positioned within, upstream or downstream of the coding sequence. In addition, the expression vectors can include additional regulatory sequences from the 3'-untranslated region of plant genes, e.g., a 3' terminator region to increase mRNA stability of the mRNA, such as the PI-II terminator region of potato or the octopine or nopaline synthase 3' terminator regions.

Additional Expression Elements

Specific initiation signals can aid in efficient translation of coding sequences. These signals can include, e.g., the ATG initiation codon and adjacent sequences. In cases where a coding sequence, its initiation codon and upstream sequences are inserted into the appropriate expression vector, no additional translational control signals may be needed. However, in cases where only coding sequence (e.g., a mature protein coding sequence), or a portion thereof, is inserted, exogenous transcriptional control signals including the ATG initiation codon can be separately provided. The initiation codon is provided in the correct reading frame to facilitate transcription. Exogenous transcriptional elements and initiation codons can be of various origins, both natural and synthetic. The efficiency of expression can be enhanced by the inclusion of enhancers appropriate to the cell system in use.

Expression Hosts

The present invention also relates to host cells which are transduced with vectors of the invention, and the production of polypeptides of the invention (including fragments thereof) by recombinant techniques. Host cells are genetically engineered (i.e., nucleic acids are introduced, e.g., transduced, transformed or transfected) with the vectors of this invention, which may be, for example, a cloning vector or an expression vector comprising the relevant nucleic acids herein. The

vector is optionally a plasmid, a viral particle, a phage, a naked nucleic acid, *etc.* The engineered host cells can be cultured in conventional nutrient media modified as appropriate for activating promoters, selecting transformants, or amplifying the relevant gene. The culture conditions, such as temperature, pH and the like, are those previously used with the host cell selected for expression, and will be apparent to those skilled in the art and in the references cited herein, including, Sambrook and Ausubel.

The host cell can be a eukaryotic cell, such as a yeast cell, or a plant cell, or the host cell can be a prokaryotic cell, such as a bacterial cell. Plant protoplasts are also suitable for some applications. For example, the DNA fragments are introduced into plant tissues, cultured plant cells or plant protoplasts by standard methods including electroporation (Fromm et al., (1985) Proc. Natl. Acad. Sci. USA 82, 5824, infection by viral vectors such as cauliflower mosaic virus (CaMV) (Hohn et al., (1982) Molecular Biology of Plant Tumors, (Academic Press, New York) pp. 549-560; US 4,407,956), high velocity ballistic penetration by small particles with the nucleic acid either within the matrix of small beads or particles, or on the surface (Klein et al., (1987) Nature 327, 70-73), use of pollen as vector (WO 85/01856), or use of *Agrobacterium tumefaciens* or *A. rhizogenes* carrying a T-DNA plasmid in which DNA fragments are cloned. The T-DNA plasmid is transmitted to plant cells upon infection by *Agrobacterium tumefaciens*, and a portion is stably integrated into the plant genome (Horsch et al. (1984) Science 233:496-498; Fraley et al. (1983) Proc. Natl. Acad. Sci. USA 80, 4803).

The cell can include a nucleic acid of the invention which encodes a polypeptide, wherein the cells expresses a polypeptide of the invention. The cell can also include vector sequences, or the like. Furthermore, cells and transgenic plants that include any polypeptide or nucleic acid above or throughout this specification, e.g., produced by transduction of a vector of the invention, are an additional feature of the invention.

For long-term, high-yield production of recombinant proteins, stable expression can be used. Host cells transformed with a nucleotide sequence encoding a polypeptide of the invention are optionally cultured under conditions suitable for the

expression and recovery of the encoded protein from cell culture. The protein or fragment thereof produced by a recombinant cell may be secreted, membrane-bound, or contained intracellularly, depending on the sequence and/or the vector used. As will be understood by those of skill in the art, expression vectors containing polynucleotides encoding mature proteins of the invention can be designed with signal sequences which direct secretion of the mature polypeptides through a prokaryotic or eukaryotic cell membrane.

XI. Modified Amino Acid Residues

Polypeptides of the invention may contain one or more modified amino acid residues. The presence of modified amino acids may be advantageous in, for example, increasing polypeptide half-life, reducing polypeptide antigenicity or toxicity, increasing polypeptide storage stability, or the like. Amino acid residue(s) are modified, for example, co-translationally or post-translationally during recombinant production or modified by synthetic or chemical means.

Non-limiting examples of a modified amino acid residue include incorporation or other use of acetylated amino acids, glycosylated amino acids, sulfated amino acids, prenylated (e.g., farnesylated, geranylgeranylated) amino acids, PEG modified (e.g., "PEGylated") amino acids, biotinylated amino acids, carboxylated amino acids, phosphorylated amino acids, etc. References adequate to guide one of skill in the modification of amino acid residues are replete throughout the literature.

The modified amino acid residues may prevent or increase affinity of the polypeptide for another molecule, including, but not limited to, polynucleotide, proteins, carbohydrates, lipids and lipid derivatives, and other organic or synthetic compounds.

XII. Identification of Additional Factors

A transcription factor provided by the present invention can also be used to identify additional endogenous or exogenous molecules that can affect a phenotype or trait of interest. On the one hand, such molecules include organic (small or large molecules) and/or inorganic compounds that affect expression of (i.e., regulate) a particular transcription factor. Alternatively, such molecules include endogenous

molecules that are acted upon either at a transcriptional level by a transcription factor of the invention to modify a phenotype as desired. For example, the transcription factors can be employed to identify one or more downstream gene with which is subject to a regulatory effect of the transcription factor. In one approach, a transcription factor or transcription factor homologue of the invention is expressed in a host cell, e.g., a transgenic plant cell, tissue or explant, and expression products, either RNA or protein, of likely or random targets are monitored, e.g., by hybridization to a microarray of nucleic acid probes corresponding to genes expressed in a tissue or cell type of interest, by two-dimensional gel electrophoresis of protein products, or by any other method known in the art for assessing expression of gene products at the level of RNA or protein. Alternatively, a transcription factor of the invention can be used to identify promoter sequences (i.e., binding sites) involved in the regulation of a downstream target. After identifying a promoter sequence, interactions between the transcription factor and the promoter sequence can be modified by changing specific nucleotides in the promoter sequence or specific amino acids in the transcription factor that interact with the promoter sequence to alter a plant trait. Typically, transcription factor DNA-binding sites are identified by gel shift assays. After identifying the promoter regions, the promoter region sequences can be employed in double-stranded DNA arrays to identify molecules that affect the interactions of the transcription factors with their promoters (Bulyk et al. (1999) Nature Biotechnology 17:573-577).

The identified transcription factors are also useful to identify proteins that modify the activity of the transcription factor. Such modification can occur by covalent modification, such as by phosphorylation, or by protein-protein (homo or heteropolymer) interactions. Any method suitable for detecting protein-protein interactions can be employed. Among the methods that can be employed are co-immunoprecipitation, cross-linking and co-purification through gradients or chromatographic columns, and the two-hybrid yeast system.

The two-hybrid system detects protein interactions in vivo and is described in Chien et al. ((1991), Proc. Natl. Acad. Sci. USA 88:9578-9582) and is commercially available from Clontech (Palo Alto, Calif.). In such a system, plasmids are constructed that encode two hybrid proteins: one consists of the DNA-binding domain

of a transcription activator protein fused to the TF polypeptide and the other consists of the transcription activator protein's activation domain fused to an unknown protein that is encoded by a cDNA that has been recombined into the plasmid as part of a cDNA library. The DNA-binding domain fusion plasmid and the cDNA library are transformed into a strain of the yeast *Saccharomyces cerevisiae* that contains a reporter gene (e.g., lacZ) whose regulatory region contains the transcription activator's binding site. Either hybrid protein alone cannot activate transcription of the reporter gene. Interaction of the two hybrid proteins reconstitutes the functional activator protein and results in expression of the reporter gene, which is detected by an assay for the reporter gene product. Then, the library plasmids responsible for reporter gene expression are isolated and sequenced to identify the proteins encoded by the library plasmids. After identifying proteins that interact with the transcription factors, assays for compounds that interfere with the TF protein-protein interactions can be preformed.

XIII. Identification of Modulators

In addition to the intracellular molecules described above, extracellular molecules that alter activity or expression of a transcription factor, either directly or indirectly, can be identified. For example, the methods can entail first placing a candidate molecule in contact with a plant or plant cell. The molecule can be introduced by topical administration, such as spraying or soaking of a plant, and then the molecule's effect on the expression or activity of the TF polypeptide or the expression of the polynucleotide monitored. Changes in the expression of the TF polypeptide can be monitored by use of polyclonal or monoclonal antibodies, gel electrophoresis or the like. Changes in the expression of the corresponding polynucleotide sequence can be detected by use of microarrays, Northern, quantitative PCR, or any other technique for monitoring changes in mRNA expression. These techniques are exemplified in Ausubel et al. (eds) Current Protocols in Molecular Biology, John Wiley & Sons (1998, and supplements through 2001). Such changes in the expression levels can be correlated with modified plant traits and thus identified molecules can be useful for soaking or spraying on fruit, vegetable and grain crops to modify traits in plants.

Essentially any available composition can be tested for modulatory activity of expression or activity of any nucleic acid or polypeptide herein. Thus, available libraries of compounds such as chemicals, polypeptides, nucleic acids and the like can be tested for modulatory activity. Often, potential modulator compounds can be dissolved in aqueous or organic (e.g., DMSO-based) solutions for easy delivery to the cell or plant of interest in which the activity of the modulator is to be tested. Optionally, the assays are designed to screen large modulator composition libraries by automating the assay steps and providing compounds from any convenient source to assays, which are typically run in parallel (e.g., in microtiter formats on microtiter plates in robotic assays).

In one embodiment, high throughput screening methods involve providing a combinatorial library containing a large number of potential compounds (potential modulator compounds). Such "combinatorial chemical libraries" are then screened in one or more assays, as described herein, to identify those library members (particular chemical species or subclasses) that display a desired characteristic activity. The compounds thus identified can serve as target compounds.

A combinatorial chemical library can be, e.g., a collection of diverse chemical compounds generated by chemical synthesis or biological synthesis. For example, a combinatorial chemical library such as a polypeptide library is formed by combining a set of chemical building blocks (e.g., in one example, amino acids) in every possible way for a given compound length (i.e., the number of amino acids in a polypeptide compound of a set length). Exemplary libraries include peptide libraries, nucleic acid libraries, antibody libraries (see, e.g., Vaughn et al. (1996) Nature Biotechnology, 14(3):309-314 and PCT/US96/10287), carbohydrate libraries (see, e.g., Liang et al. Science (1996) 274:1520-1522 and U.S. Patent 5,593,853), peptide nucleic acid libraries (see, e.g., U.S. Patent 5,539,083), and small organic molecule libraries (see, e.g., benzodiazepines, Baum C&EN Jan 18, page 33 (1993); isoprenoids, U.S. Patent 5,569,588; thiazolidinones and metathiazanones, U.S. Patent 5,549,974; pyrrolidines, U.S. Patents 5,525,735 and 5,519,134; morpholino compounds, U.S. Patent 5,506,337) and the like.

Preparation and screening of combinatorial or other libraries is well known to those of skill in the art. Such combinatorial chemical libraries include, but are not limited to, peptide libraries (see, e.g., U.S. Patent 5,010,175; Furka, (1991) Int. J. Pept. Prot. Res. 37:487-493; and Houghton et al. (1991) Nature 354:84-88). Other chemistries for generating chemical diversity libraries can also be used.

In addition, as noted, compound screening equipment for high-throughput screening is generally available, e.g., using any of a number of well known robotic systems that have also been developed for solution phase chemistries useful in assay systems. These systems include automated workstations including an automated synthesis apparatus and robotic systems utilizing robotic arms. Any of the above devices are suitable for use with the present invention, e.g., for high-throughput screening of potential modulators. The nature and implementation of modifications to these devices (if any) so that they can operate as discussed herein will be apparent to persons skilled in the relevant art.

Indeed, entire high throughput screening systems are commercially available. These systems typically automate entire procedures including all sample and reagent pipetting, liquid dispensing, timed incubations, and final readings of the microplate in detector(s) appropriate for the assay. These configurable systems provide high throughput and rapid start up as well as a high degree of flexibility and customization. Similarly, microfluidic implementations of screening are also commercially available.

The manufacturers of such systems provide detailed protocols the various high throughput. Thus, for example, Zymark Corp. provides technical bulletins describing screening systems for detecting the modulation of gene transcription, ligand binding, and the like. The integrated systems herein, in addition to providing for sequence alignment and, optionally, synthesis of relevant nucleic acids, can include such screening apparatus to identify modulators that have an effect on one or more polynucleotides or polypeptides according to the present invention.

In some assays it is desirable to have positive controls to ensure that the components of the assays are working properly. At least two types of positive controls are appropriate. That is, known transcriptional activators or inhibitors can be

incubated with cells/plants/ etc. in one sample of the assay, and the resulting increase/decrease in transcription can be detected by measuring the resulting increase in RNA/ protein expression, etc., according to the methods herein. It will be appreciated that modulators can also be combined with transcriptional activators or inhibitors to find modulators that inhibit transcriptional activation or transcriptional repression. Either expression of the nucleic acids and proteins herein or any additional nucleic acids or proteins activated by the nucleic acids or proteins herein, or both, can be monitored.

In an embodiment, the invention provides a method for identifying compositions that modulate the activity or expression of a polynucleotide or polypeptide of the invention. For example, a test compound, whether a small or large molecule, is placed in contact with a cell, plant (or plant tissue or explant), or composition comprising the polynucleotide or polypeptide of interest and a resulting effect on the cell, plant, (or tissue or explant) or composition is evaluated by monitoring, either directly or indirectly, one or more of: expression level of the polynucleotide or polypeptide, activity (or modulation of the activity) of the polynucleotide or polypeptide. In some cases, an alteration in a plant phenotype can be detected following contact of a plant (or plant cell, or tissue or explant) with the putative modulator, e.g., by modulation of expression or activity of a polynucleotide or polypeptide of the invention. Modulation of expression or activity of a polynucleotide or polypeptide of the invention may also be caused by molecular elements in a signal transduction second messenger pathway and such modulation can affect similar elements in the same or another signal transduction second messenger pathway.

XIV. Subsequences

Also contemplated are uses of polynucleotides, also referred to herein as oligonucleotides, typically having at least 12 bases, preferably at least 15, more preferably at least 20, 30, or 50 bases, which hybridize under at least highly stringent (or ultra-high stringent or ultra-ultra-high stringent conditions) conditions to a polynucleotide sequence described above. The polynucleotides may be used as probes, primers, sense and antisense agents, and the like, according to methods as noted *supra*.

Subsequences of the polynucleotides of the invention, including polynucleotide fragments and oligonucleotides are useful as nucleic acid probes and primers. An oligonucleotide suitable for use as a probe or primer is at least about 15 nucleotides in length, more often at least about 18 nucleotides, often at least about 21 nucleotides, frequently at least about 30 nucleotides, or about 40 nucleotides, or more in length. A nucleic acid probe is useful in hybridization protocols, e.g., to identify additional polypeptide homologues of the invention, including protocols for microarray experiments. Primers can be annealed to a complementary target DNA strand by nucleic acid hybridization to form a hybrid between the primer and the target DNA strand, and then extended along the target DNA strand by a DNA polymerase enzyme. Primer pairs can be used for amplification of a nucleic acid sequence, e.g., by the polymerase chain reaction (PCR) or other nucleic-acid amplification methods. See Sambrook and Ausubel, *supra*.

In addition, the invention includes an isolated or recombinant polypeptide including a subsequence of at least about 15 contiguous amino acids encoded by the recombinant or isolated polynucleotides of the invention. For example, such polypeptides, or domains or fragments thereof, can be used as immunogens, e.g., to produce antibodies specific for the polypeptide sequence, or as probes for detecting a sequence of interest. A subsequence can range in size from about 15 amino acids in length up to and including the full length of the polypeptide.

To be encompassed by the present invention, an expressed polypeptide which comprises such a polypeptide subsequence performs at least one biological function of the intact polypeptide in substantially the same manner, or to a similar extent, as does the intact polypeptide. For example, a polypeptide fragment can comprise a recognizable structural motif or functional domain such as a DNA binding domain that binds to a specific DNA promoter region, an activation domain or a domain for protein-protein interactions.

XV. Production of Transgenic Plants

Modification of Traits

The polynucleotides of the invention are favorably employed to produce transgenic plants with various traits, or characteristics, that have been modified in a desirable manner, e.g., to improve the seed characteristics of a plant. For example, alteration of expression levels or patterns (e.g., spatial or temporal expression patterns) of one or more of the transcription factors (or transcription factor homologues) of the invention, as compared with the levels of the same protein found in a wild type plant, can be used to modify a plant's traits. An illustrative example of trait modification, improved characteristics, by altering expression levels of a particular transcription factor is described further in the Examples and the Sequence Listing.

Arabidopsis as a model system

Arabidopsis thaliana is the object of rapidly growing attention as a model for genetics and metabolism in plants. *Arabidopsis* has a small genome, and well documented studies are available. It is easy to grow in large numbers and mutants defining important genetically controlled mechanisms are either available, or can readily be obtained. Various methods to introduce and express isolated homologous genes are available (see Koncz, et al., eds. *Methods in Arabidopsis Research*, et al. (1992), World Scientific, New Jersey, New Jersey, in "Preface"). Because of its small size, short life cycle, obligate autogamy and high fertility, *Arabidopsis* is also a choice organism for the isolation of mutants and studies in morphogenetic and development pathways, and control of these pathways by transcription factors (Koncz, *supra*, p. 72). A number of studies introducing transcription factors into *A. thaliana* have demonstrated the utility of this plant for understanding the mechanisms of gene regulation and trait alteration in plants. See, for example, Koncz, *supra*, and U.S. Patent Number 6,417,428).

Arabidopsis genes in transgenic plants.

Expression of genes which encode transcription factors modify expression of endogenous genes, polynucleotides, and proteins are well known in the art. In addition, transgenic plants comprising isolated polynucleotides encoding transcription factors may also modify expression of endogenous genes, polynucleotides, and

proteins. Examples include Peng et al. (1997, *Genes and Development* 11:3194-3205) and Peng et al. (1999, *Nature*, 400:256-261). In addition, many others have demonstrated that an *Arabidopsis* transcription factor expressed in an exogenous plant species elicits the same or very similar phenotypic response. See, for example, Fu et al. (2001, *Plant Cell* 13:1791-1802); Nandi et al. (2000, *Curr. Biol.* 10:215-218); Coupland (1995, *Nature* 377:482-483); and Weigel and Nilsson (1995, *Nature* 377:482-500).

Homologous genes introduced into transgenic plants.

Homologous genes that may be derived from any plant, or from any source whether natural, synthetic, semi-synthetic or recombinant, and that share significant sequence identity or similarity to those provided by the present invention, may be introduced into plants, for example, crop plants, to confer desirable or improved traits. Consequently, transgenic plants may be produced that comprise a recombinant expression vector or cassette with a promoter operably linked to one or more sequences homologous to presently disclosed sequences. The promoter may be, for example, a plant or viral promoter.

The invention thus provides for methods for preparing transgenic plants, and for modifying plant traits. These methods include introducing into a plant a recombinant expression vector or cassette comprising a functional promoter operably linked to one or more sequences homologous to presently disclosed sequences. Plants and kits for producing these plants that result from the application of these methods are also encompassed by the present invention.

The complete descriptions of the traits associated with each polynucleotide of the invention is fully disclosed in Table 4, Table 5, and Table 6.

Table 4

Polynucleotide SEQ ID NO:	GID No.	Trait	Category	Family	Comment	Polypeptide SEQ ID NO:	Conserved domains
1	G1275	Architecture; size	Dev and morph	WRKY	Reduced apical dominance; small plant	2	(113-169)
3	G1411	Architecture	Dev and morph	AP2	Loss of apical dominance	4	(87-154)
5	G1488	Architecture; light response; size; seed protein content	Dev and morph; seed biochemistry	GATAZn	Reduced apical dominance, shorter stems; constitutive photomorphogenesis; reduced size; altered seed protein content	6	(221-246)
7	G1499	Architecture; flower; morphology; other	Dev and morph	HLH/MYC	Altered plant architecture; altered floral organ identity and development; dark green color	8	(118-181)
9	G1543	Architecture; flower; morphology; other; seed oil	Dev and morph; seed biochemistry	HB	Altered plant architecture; altered carpal shape; dark green color; decreased seed oil	10	(135-195)
11	G1635	Architecture; morphology; other; fertility	Dev and morph	MYB-related	Reduced apical dominance; pale green, smaller plants; reduced fertility	12	(44-104)
13	G1794	Architecture; light response; seed oil and protein content	Dev and morph; seed biochemistry	AP2	Altered plant architecture; constitutive photomorphogenesis; altered seed oil and protein content	14	(182-248)
15	G1839	Architecture; size	Dev and morph	AP2	Altered plant architecture; reduced size	16	(118-184)
17	G2108	Architecture	Dev and morph	AP2	Altered inflorescence structure	18	(18-85)
19	G2291	Architecture; flowering time	Dev and morph; flowering time	AP2	Altered plant architecture; late flowering	20	(TBD)
21	G2452	Architecture; leaf	Dev and morph	MYB-related	Reduced apical dominance; pale green color	22	(27-213)
23	G2509	Architecture; seed oil and protein content	Dev and morph; seed biochemistry	AP2	Reduced apical dominance; altered seed oil and protein content	24	(89-156)
25	G390	Architecture	Dev and morph	HB	Altered shoot development	26	(18-81)
27	G391	Architecture	Dev and morph	HB	Altered shoot development	28	(25-85)
29	G438	Architecture; stem	Dev and morph	HB	Reduced branching; reduced lignin	30	(22-85)

Table 4

31	G47	Architecture; stem; flowering time; altered seed oil content	Dev and morph; flowering time; seed biochemistry	AP2	Altered architecture and inflorescence development, structure of vascular tissues; late flowering; altered seed oil content	32	(11-80)
33	G559	Architecture; fertility	Dev and morph	bZIP	Loss of apical dominance; reduced fertility	34	(203-264)
35	G568	Architecture; flowering time	Dev and morph	bZIP	Altered branching; late flowering	36	(215-265)
37	G580	Architecture; flower	Dev and morph	bZIP	Altered inflorescences; altered flower development	38	(162-218)
39	G615	Architecture; fertility	Dev and morph	TEO	Altered plant architecture; little or no pollen production, poor filament elongation	40	(88-147)
41	G732	Architecture; flower; seed oil and protein	Dev and morph; seed biochemistry	bZIP	Reduced apical dominance; abnormal flowers; altered seed oil and protein content	42	(31-91)
43	G988	Architecture; fertility; flower; stem; seed oil and protein content	Dev and morph; seed biochemistry	SCR	Reduced lateral branching; reduced fertility; enlarged floral organs, short pedicels; thicker stem, altered distribution of vascular bundles; altered seed oil and protein content	44	(178-195)
45	G1519	Embryo lethal	Dev and morph	RING/C3HC4	Embryo lethal	46	(327-364)
47	G374	Embryo lethal	Dev and morph	Z-ZPF	Embryo lethal	48	(35-67, 245-277)
49	G877	Embryo lethal	Dev and morph	WRKY	Embryo lethal	50	(272-328, 487-603)
51	G1000	Fertility; size; flower; stem	Dev and morph	MYB-(R1)R2R3	Reduced fertility; small plant; reduced or absent petals and sepals; reduced inflorescence, stem elongation	52	(14-117)
53	G1067	Fertility; leaf	Dev and morph	AT-hook	Reduced fertility; altered leaf shape; small plant	54	(86-93)
55	G1075	Fertility; flower; leaf; size	Dev and morph	AT-hook	Reduced fertility; reduced or absent petals, sepals and stamens; altered leaf shape; small plant	56	(78-85)
57	G1266	Fertility; size	Dev and morph	AP2	Reduced fertility; small plant	58	(79-147)
59	G1311	Fertility; size	Dev and morph	MYB-(R1)R2R3	Reduced fertility; small plant	60	(11-112)
61	G1321	Fertility; flower	Dev and morph	MYB-(R1)R2R3	Poor fertility; altered flower morphology	62	(4-106)
63	G1326	Fertility; flower; size	Dev and morph	MYB-(R1)R2R3	Reduced fertility; petals and sepals are smaller; small plant	64	(18-121)
65	G1367	Fertility; size	Dev and morph	AT-hook	Reduced fertility; reduced size	66	(179-201, 262-285, 298-319, 335-357)
67	G1386	Fertility; size; seed oil and protein content	Dev and morph; seed biochemistry	AP2	Reduced fertility; reduced size; altered seed oil and protein content	68	(TBD)

Table 4

69	G1421	Fertility; size; seed oil content	Dev and morph; seed biochemistry	AP2	Reduced fertility; small plant; altered seed oil content	70	(74-151)
71	G1453	Fertility; morphology; other	Dev and morph	NAC	Reduced fertility; altered inflorescence development	72	(13-160)
73	G1560	Fertility; flower; size	Dev and morph	HS	Reduced fertility; altered flower development; reduced size	74	(62-151)
75	G1594	Fertility; leaf; seed	Dev and morph	HB	Reduced fertility; altered leaf shape and development; large pale seed	76	(343-308)
77	G1750	Fertility; size; seed oil content	Dev and morph; seed biochemistry	AP2	Reduced fertility; reduced size; increased seed oil content	78	(107-173)
79	G1947	Fertility; flower; seed protein content	Dev and morph; seed biochemistry	HS	Reduced fertility; extended period of flowering; altered seed protein content	80	(37-120)
81	G2011	Fertility; size; seed oil and protein content	Dev and morph; seed biochemistry	HS	Reduced fertility; reduced size; altered seed oil and protein content	82	(56-147)
83	G2094	Fertility; leaf; size	Dev and morph	GATA/Zn	Reduced fertility; altered leaf development; reduced size	84	(43-68)
85	G2113	Fertility; leaf; seed protein content	Dev and morph; seed biochemistry	AP2	Reduced fertility; long petioles, altered orientation; altered seed protein content	86	(TBD)
87	G2115	Fertility; size	Dev and morph	AP2	Reduced fertility; reduced size	88	(46-115)
89	G2130	Fertility; size; senescence	Dev and morph	AP2	Reduced fertility; reduced size; early senescence	90	(93-160)
91	G2147	Fertility; size	Dev and morph	HLH/MYC	Reduced fertility; reduced size	92	(160-234)
93	G2156	Fertility; size; seed protein content	Dev and morph; seed biochemistry	AT-hook	Reduced fertility; reduced size; altered seed protein content	94	(66-86)
95	G2294	Fertility; size	Dev and morph	AP2	Reduced fertility; reduced size	96	(32-102)
97	G2510	Fertility; size	Dev and morph	AP2	Reduced fertility; reduced size	98	(41-108)
99	G2893	Fertility; flower; size	Dev and morph	MYB-(R1)R2R3	Reduced fertility; altered flower development; reduced size	100	(19-120)
101	G340	Fertility; size	Dev and morph	Z-C3H	Reduced fertility, size	102	(37-154)
103	G39	Fertility; size	Dev and morph	AP2	Reduced fertility, small plant	104	(24-90)
105	G439	Fertility; size	Dev and morph	AP2	Reduced fertility, small plant	106	(110-177)
107	G470	Fertility	Dev and morph	ARF	Short stamen filaments	108	(61-393)

Table 4

109	G652	Fertility; seed; flower; size; seed oil content	Dev and morph; seed biochemistry	Z-CLDSH	Reduced fertility; irregular shaped seed; altered flower development; reduced size, slow growth; altered seed oil content	110	(28-49, 137-151, 182-196)
111	G671	Fertility; flower; leaf; size; stem	Dev and morph; biochemistry	MYB-(R1)R2R3	Reduced fertility; reduced petal abscission; altered leaf shape; small plant; altered inflorescence stem structure	112	(15-115)
113	G779	Fertility; flower	Dev and morph; biochemistry	HLH/MYC	Reduced fertility, homeotic transformations	114	(126-182)
115	G962	Fertility; size	Dev and morph; biochemistry	NAC	Reduced fertility; small plant	116	(53-175)
117	G977	Fertility; leaf; morphology; other; size	Dev and morph; biochemistry	AP2	Reduced fertility; altered leaf shape; dark green; small plant	118	(5-72)
119	G1063	Flower; leaf; inflorescence; seed oil and protein content	Dev and morph; seed biochemistry	HLH/MYC	Altered flower development, ectopic carpel tissue; altered leaf shape, dark green color; altered inflorescence development; altered seed oil and protein content	120	(131-182)
121	G1140	Flower	Dev and morph; biochemistry	MADS	Altered flower development	122	(2-57)
123	G1425	Flower	Dev and morph; biochemistry	NAC	Altered flower and inflorescence development	124	(20-173)
125	G1449	Flower	Dev and morph; biochemistry	IAA	Altered flower structure	126	(48-53, 74-107, 122-152)
127	G1897	Flower; leaf; seed protein content	Dev and morph; seed biochemistry	Z-Dof	Altered flower development; altered leaf development; altered seed protein content	128	(34-62)
129	G2143	Flower; leaf; inflorescence	Dev and morph; biochemistry	HLH/MYC	Altered flower development, ectopic carpel tissue; altered leaf shape, dark green color; altered inflorescence development	130	(128-179)
131	G2535	Flower; seed protein content	Dev and morph; seed biochemistry	NAC	Altered flower development; altered seed protein content	132	(11-114)
133	G2557	Flower; leaf	Dev and morph; biochemistry	HLH/MYC	Altered flower development, ectopic carpel tissue; altered leaf shape, dark green color	134	(278-328)
135	G259	Flower; leaf	Dev and morph; biochemistry	HS	Altered flower development; altered leaf development	136	(27-131)
137	G353	Flower; leaf; size; seed protein content	Dev and morph; seed biochemistry	Z-C2H2	Short pedicels, downward pointing siliques; altered leaf development; reduced size; altered seed protein content	138	(41-61, 84-104)
139	G354	Flower; light response; size	Dev and morph; biochemistry	Z-C2H2	Short pedicels, downward pointing siliques; constitutive morphogenesis; reduced size	140	(42-62, 88-109)
141	G638	Flower; morphology; other	Dev and morph; biochemistry	TH	Altered flower development; multiple developmental defects	142	(119-206)

Table 4

143	G869	Flower; morphology; other; seed oil	Dev and morph; seed biochemistry	AP2	Abnormal anther development; small and spindly plant; altered seed fatty acids	144	(109-177)
145	G1645	Inflorescence; leaf	Dev and morph	MYB-(R1)R2R3	Altered inflorescence structure; altered leaf development	146	(90-210)
147	G1038	Leaf	Dev and morph	GARP	Altered leaf shape	148	(198-247)
149	G1073	Leaf; size; flowering time	Dev and morph; flowering time	AT-hook	Serrated leaves; increased plant size; flowering appears to be slightly delayed	150	(33-42, 78-175)
151	G1146	Leaf	Dev and morph	PAZ	Altered leaf development	152	(886-896)
153	G1267	Leaf; size	Dev and morph	WRKY	Dark green shiny leaves; small plant	154	(70-127)
155	G1269	Leaf	Dev and morph	MYB-related	Long petioles, upturned leaves	156	(27-83)
157	G1452	Leaf; trichome; flowering time	Dev and morph; flowering time	NAC	Altered leaf shape, dark green color; reduced trichome density; late flowering	158	(30-177)
159	G1494	Leaf; size; light response; seed	Dev and morph	HLH/MYC	Pale green leaves; altered leaf shape; reduced size; long hypocotyls; large, pale seeds	160	(261-311)
161	G1548	Leaf	Dev and morph	HB	Altered leaf development	162	(17-77)
163	G1574	Leaf	Dev and morph	SWI/SNF	Altered leaf development	164	(28-350)
165	G1586	Leaf; size	Dev and morph	HB	Narrow leaves; small plants	166	(21-81)
167	G1786	Leaf; light response; size	Dev and morph	MYB-(R1)R2R3	Dark green, small leaves with short petioles; photomorphogenesis in the dark; small plant	168	(TBD)
169	G1792	Leaf; seed oil and protein content	Dev and morph; seed biochemistry	AP2	Dark green, shiny leaves; altered seed oil and protein content	170	(17-85)
171	G1865	Leaf; seed oil and protein content	Dev and morph; seed biochemistry	GRF-like	Altered leaf development; altered seed oil and protein content	172	(124-149)
173	G1886	Leaf; size	Dev and morph	Z-Dof	Chlorotic patches in leaves; reduced size	174	(17-59)
175	G1933	Leaf; size; seed protein content	Dev and morph; seed biochemistry	WRKY	Altered leaf development; reduced size; altered seed protein content	176	(205-263, 344-404)
177	G2059	Leaf; seed oil and protein content	Dev and morph; seed biochemistry	AP2	Smaller, curled leaves; altered seed oil, protein content	178	(184-254)
179	G2105	Leaf; seed oil and protein content	Dev and morph	TH	Alterations in leaf surface; large, pale seeds	180	(100-153)
181	G2117	Leaf; seed oil and protein content	Dev and morph; seed biochemistry	bZIP	Small, dark green leaves; altered seed oil and protein content	182	(46-106)

Table 4

183	G2124	Leaf; seed protein content	Dev and morph; biochemistry	TEO		Altered leaf development; altered seed protein content	184	(75-132)
185	G2140	Leaf; root	Dev and morph	HLH/MYC		Altered leaf development; short roots	186	(167-242)
187	G2144	Leaf; light response; size; seed oil content	Dev and morph; biochemistry	HLH/MYC		Pale green leaves, altered leaf shape; long hypocotyls; reduced size; altered seed oil content	188	(203-283)
189	G2431	Leaf	Dev and morph	GARP		Dark green leaves; reduced size	190	(38-88)
191	G2465	Morphology; other; leaf	Dev and morph	GARP		Slowed development; altered leaf color and shape	192	(219-269)
193	G2583	Leaf; seed oil and protein content	Dev and morph; biochemistry	AP2		Glossy, shiny leaves; altered seed oil and protein content	194	(4-71)
195	G2724	Leaf	Dev and morph	MYB-(R1)R2R3		Dark green leaves	196	(7-113)
197	G377	Morphology; other	Dev and morph	RING/C3H2C3		Altered leaf development; slow growth	198	(85-128)
199	G428	Leaf	Dev and morph	HB		Altered leaf shape	200	(229-292)
201	G447	Leaf; morphology; other; size	Dev and morph	ARF		Dark green leaves; altered cotyledon shape; reduced size	202	(22-356)
203	G464	Leaf	Dev and morph	IAA		Altered leaf shape	204	(20-28, 71-82, 126-142, 187-224)
205	G557	Leaf; size	Dev and morph	bZIP		Dark green color; small plant	206	(90-150)
207	G577	Leaf	Dev and morph	BZIP2		Reduced size, increased anthocyanins	208	(TBD)
209	G674	Leaf; size	Dev and morph	MYB-(R1)R2R3		Dark green leaves, upwardly oriented; reduced size	210	(20-120)
211	G736	Leaf; flowering time	Dev and morph; flowering time	Z-Dof		Altered leaf shape; later flowering	212	(54-111)
213	G903	Leaf	Dev and morph	Z-C2H2		Altered leaf morphology	214	(68-92)
215	G917	Leaf; seed oil and protein content	Dev and morph; biochemistry	MADS		Altered leaf development; altered seed oil and protein content	216	(2-57)
217	G921	Leaf	Dev and morph	WRKY		Serrated leaves	218	(146-203)
219	G922	Leaf; size	Dev and morph	SCR		Altered development, dark green color; reduced size	220	(225-242)
221	G932	Leaf; size	Dev and morph	MYB-(R1)R2R3		Altered development, dark green color; reduced size	222	(12-118)
223	G599	Leaf; size	Dev and morph	DBP		Altered leaf shape; small plant	224	(187-219, 264-300)
225	G804	Leaf; size	Dev and morph	PCF		Altered leaf shape, small plant	226	(54-117)

Table 4

227	G1062	Light response; morphology; other; seed size	Dev and morph	HLH/MYC	Constitutive photomorphogenesis; altered seed shape	228	(308-359)
229	G1322	Light response; size	Dev and morph	MYB-(R1)R2R3	Photomorphogenesis in the dark; reduced size	230	(26-130)
231	G1331	Light response; morphology; other; seed oil and protein content	Dev and morph; seed biochemistry	MYB-(R1)R2R3	Constitutive photomorphogenesis; multiple developmental alterations; altered seed oil and protein content	232	(8-109)
233	G1521	Light response	Dev and morph	RING/C3HC4	Constitutive photomorphogenesis	234	(39-80)
235	G183	Light response; seed protein content	Dev and morph; seed biochemistry	WRKY	Constitutive photomorphogenesis; altered seed protein content	236	(307-363)
237	G2555	Light response	Dev and morph	HLH/MYC	Constitutive photomorphogenesis	238	(175-245)
239	G375	Light response	Dev and morph	Z-Dof	Upward pointing leaves	240	(75-103)
241	G1007	Morphology; other	Dev and morph	AP2	Multiple developmental alterations	242	(TBD)
243	G1010	Morphology; other	Dev and morph	ABI3/VP-1	Multiple developmental alterations	244	(33-122)
245	G1014	Morphology; other; trichome	Dev and morph	ABI3/VP-1	Multiple developmental defects; reduced trichomes	246	(90-172)
247	G1035	Morphology; other	Dev and morph	bZIP	Multiple developmental alterations	248	(39-91)
249	G1046	Morphology; other	Dev and morph	bZIP	Multiple developmental alterations	250	(79-138)
251	G1049	Morphology; other; seed protein content	Dev and morph; seed biochemistry	bZIP	Multiple developmental alterations; altered seed protein content	252	(77-132)
253	G1069	Morphology; other; seed oil content	Dev and morph; seed biochemistry	AT-hook	Multiple developmental alterations; altered seed oil content	254	(67-74)
255	G1070	Morphology; other	Dev and morph	AT-hook	Several developmental defects	256	(98-120)
257	G1076	Morphology; other	Dev and morph	AT-hook	Lethal when overexpressed	258	(82-89)
259	G1089	Morphology; other	Dev and morph	BZIPT2	Developmental defects at seedling stage	260	(425-500)
261	G1093	Morphology; other	Dev and morph	RING/C3H2C3	Multiple morphological alterations	262	(105-148)

Table 4

263	G1127	Morphology: other	Dev and morph	AT-hook	Multiple developmental alterations	264	(103-110, 155-162)
265	G1131	Morphology: other; seed protein content	Dev and morph; seed biochemistry	HLH/MYC	Multiple developmental alterations; altered seed protein content	266	(173-220)
267	G1145	Morphology: other; seed oil and protein	Dev and morph; seed biochemistry	bZIP	Multiple developmental alterations; reduced seed size, altered seed shape; altered seed oil and protein content	268	(227-270)
269	G1229	Morphology: other; seed oil and protein content	Dev and morph; seed biochemistry	HLH/MYC	Several developmental defects; altered seed oil and protein content	270	(102-160)
271	G1246	Morphology: other; seed protein content	Dev and morph; seed biochemistry	MYB-(R1)R2R3	Multiple developmental alterations; altered seed protein content	272	(27-139)
273	G1255	Morphology: other; seed	Dev and morph	Z-CO-like	Reduced apical dominance; increased seed size	274	(18-56)
275	G1304	Morphology: other	Dev and morph	MYB-(R1)R2R3	Lethal when overexpressed	276	(13-118)
277	G1318	Morphology: other	Dev and morph	MYB-(R1)R2R3	Multiple developmental alterations	278	(20-123)
279	G1320	Morphology: other	Dev and morph	MYB-(R1)R2R3	Multiple developmental alterations	280	(5-108)
281	G1330	Morphology: other	Dev and morph	MYB-(R1)R2R3	Multiple developmental alterations	282	(28-134)
283	G1352	Morphology: other	Dev and morph	Z-C2H2	Multiple developmental alterations	284	(108-129, 167-188)
285	G1354	Morphology: other	Dev and morph	NAC	Multiple developmental alterations	286	(TBD)
287	G1360	Morphology: other	Dev and morph	NAC	Lethal when overexpressed	288	(18-174)
289	G1364	Morphology: other	Dev and morph	CAAT	Lethal when overexpressed	290	(29-120)
291	G1379	Morphology: other	Dev and morph	AP2	Multiple developmental alterations	292	(18-85)
293	G1384	Morphology: other	Dev and morph	AP2	Abnormal inflorescence and flower development	294	(TBD)
295	G1399	Morphology: other	Dev and morph	AT-hook	Multiple developmental alterations	296	(86-93)

Table 4

297	G1415	Morphology: other	Dev and morph	AP2	Multiple developmental alterations	298	(TBD)
299	G1417	Morphology: other; seed oil	Dev and morph; biochemistry	WRKY	Reduced seedling germination and vigor; increase in 18:2, decrease in 18:3	300	(239-296)
301	G1442	Morphology: other	Dev and morph	GRF-like	Multiple developmental alterations	302	(172-223)
303	G1454	Morphology: other; seed oil and protein content	Dev and morph; biochemistry	NAC	Multiple developmental alterations; altered seed oil and protein content	304	(9-178)
305	G1459	Morphology: other	Dev and morph	NAC	Multiple developmental alterations	306	(10-152)
307	G1460	Morphology: other; seed protein content	Dev and morph; biochemistry	NAC	Multiple developmental alterations; altered seed protein content	308	(TBD)
309	G147	Morphology: other	Dev and morph	MADS	Multiple developmental defects	310	(2-57)
311	G1471	Morphology: other; seed oil	Dev and morph; biochemistry	Z-C2H2	Multiple developmental alterations; increased seed oil content	312	(49-70)
313	G1475	Morphology: other	Dev and morph	Z-C2H2	Multiple developmental alterations	314	(51-73)
315	G1477	Morphology: other	Dev and morph	Z-C2H2	Multiple developmental alterations	316	(29-48)
317	G1487	Morphology: other; seed oil and protein content	Dev and morph; biochemistry	GATA/Zn	Multiple developmental alterations; altered seed oil and protein content	318	(251-276)
319	G1492	Morphology: other	Dev and morph	GARP	Multiple developmental alterations	320	(34-83)
321	G1531	Morphology: other; seed; seed protein content	Dev and morph; biochemistry	RING/C3HC4	Multiple developmental alterations; pale seed; altered seed protein content	322	(41-77)
323	G1540	Morphology: other	Dev and morph	HB	Reduced cell differentiation in meristem	324	(35-98)
325	G1544	Morphology: other	Dev and morph	HB	Multiple developmental alterations	326	(64-124)

Table 4

327	G156	Morphology: other; seed	Dev and morph	MADS	Multiple developmental defects; seed color alteration	328	(2-57)
329	G1584	Morphology: other	Dev and morph	HB	Multiple developmental alterations	330	(TBD)
331	G1587	Morphology: other	Dev and morph	HB	Multiple developmental alterations	332	(61-121)
333	G1588	Morphology: other	Dev and morph	HB	Multiple developmental alterations	334	(66-124)
335	G1589	Morphology: other; seed protein content	Dev and morph; seed biochemistry	HB	Multiple developmental alterations; altered seed protein content	336	(384-448)
337	G160	Morphology: other	Dev and morph	MADS	Multiple developmental defects	338	(7-62)
339	G1636	Morphology: other	Dev and morph	MYB-related	Pale green, smaller plants	340	(100-165)
341	G1642	Morphology: other	Dev and morph	MYB-(R1)R2R3	Multiple developmental alterations	342	(TBD)
343	G1747	Morphology: other; seed protein content	Dev and morph; seed biochemistry	MYB-(R1)R2R3	Multiple developmental alterations; altered seed protein content	344	(11-114)
345	G1749	Morphology: other	Dev and morph	AP2	Multiple developmental alterations; formation of necrotic lesions	346	(84-155)
347	G1751	Morphology: other	Dev and morph	AP2	Multiple developmental alterations	348	(TBD)
349	G1752	Morphology: other	Dev and morph	AP2	Multiple developmental alterations	350	(83-151)
351	G1763	Morphology: other	Dev and morph	AP2	Lethal when overexpressed	352	(140-209)
353	G1766	Morphology: other	Dev and morph	NAC	Multiple developmental alterations	354	(10-153)
355	G1767	Morphology: other; seed oil content	Dev and morph; seed biochemistry	SCR	Multiple developmental alterations; altered seed oil content	356	(255-272)
357	G1778	Morphology: other	Dev and morph	GATA/Zn	Lethal when overexpressed	358	(94-119)
359	G1789	Morphology: other; seed protein content	Dev and morph; seed biochemistry	MYB-related	Delayed development; altered seed protein content	360	(1-50)
361	G1790	Morphology: other	Dev and morph	MYB-(R1)R2R3	Lethal when overexpressed	362	(217-316)

Table 4

363	G1791	Morphology: other	Dev and morph	AP2	Multiple developmental alterations	364	(TBD)
365	G1793	Morphology: other; seed oil	Dev and morph; biochemistry	AP2	Multiple developmental alterations; increased seed oil content	366	(179-255, 281-349)
367	G1795	Morphology: other; trichome	Dev and morph	AP2	Multiple developmental alterations; reduced trichomes	368	(12-80)
369	G1800	Morphology: other	Dev and morph	AP2	Multiple developmental alterations	370	(TBD)
371	G1806	Morphology: other	Dev and morph	bZIP	Multiple developmental alterations	372	(165-225)
373	G1811	Morphology: other	Dev and morph	ABI3/VP-1	Multiple developmental alterations	374	(TBD)
375	G182	Morphology: other	Dev and morph	WRKY	Multiple developmental alterations	376	(217-276)
377	G1835	Morphology: other	Dev and morph	GATA/Zn	Small, spindly plant	378	(224-296)
379	G1836	Morphology: other	Dev and morph	CAAT	Pale green	380	(30-164)
381	G1838	Morphology: other; seed oil content	Dev and morph; biochemistry	AP2	Multiple developmental alterations; increased seed oil content	382	(229-305, 330-400)
383	G1843	Morphology: other	Dev and morph	MADS	Multiple developmental alterations	384	(2-57)
385	G1853	Morphology: other	Dev and morph	AKR	Lethal when overexpressed	386	(entire protein)
387	G1855	Morphology: other	Dev and morph	AKR	Slow growth	388	(entire protein)
389	G187	Morphology: other	Dev and morph	WRKY	Variety of morphological alterations	390	(172-228)
391	G1881	Morphology: other	Dev and morph	Z-CO-like	Multiple developmental alterations	392	(5-28, 56-79)
393	G1882	Morphology: other	Dev and morph	Z-Dof	Lethal when overexpressed	394	(97-125)
395	G1883	Morphology: other	Dev and morph	Z-Dof	Multiple developmental alterations	396	(82-124)
397	G1884	Morphology: other	Dev and morph	Z-Dof	Multiple developmental alterations	398	(43-71)
399	G1891	Morphology: other	Dev and morph	Z-Dof	Multiple developmental alterations	400	(27-69)

Table 4

401	G1896	Morphology: other	Dev and morph	Z-Dof	Multiple developmental alterations	402	(43-85)
403	G1898	Morphology: other	Dev and morph	Z-Dof	Lethal when overexpressed	404	(31-59)
405	G1902	Morphology: other; seed oil content	Dev and morph; seed biochemistry	Z-Dof	Multiple developmental alterations; increased seed oil content	406	(31-59)
407	G1904	Morphology: other	Dev and morph	Z-Dof	Multiple developmental alterations	408	(53-95)
409	G1906	Morphology: other	Dev and morph	Z-Dof	Multiple developmental alterations	410	(19-47)
411	G1913	Morphology: other	Dev and morph	Z-Dof	Lethal when overexpressed	412	(27-55)
413	G1914	Morphology: other	Dev and morph	Z-C2H2	Multiple developmental alterations	414	(195-216, 245-266)
415	G1925	Morphology: other	Dev and morph	NAC	Multiple developmental alterations	416	(6-150)
417	G1929	Morphology: other	Dev and morph	Z-CO-like	Slow growth, delayed development	418	(31-53)
419	G1930	Morphology: other	Dev and morph	AP2	Multiple developmental alterations	420	(59-124)
421	G195	Morphology: other	Dev and morph	WRKY	Multiple developmental defects	422	(183-239)
423	G1954	Morphology: other	Dev and morph	HLH/MYC	Lethal when overexpressed	424	(187-259)
425	G1958	Morphology: other; seed protein content	Dev and morph; seed biochemistry	GARP	Reduced size and root mass in plates.; altered seed protein content	426	(230-278)
427	G196	Morphology: other; seed protein content	Dev and morph; seed biochemistry	WRKY	Multiple developmental alterations; altered seed protein content	428	(223-283)
429	G1965	Morphology: other	Dev and morph	Z-Dof	Lethal when overexpressed	430	(27-55)
431	G1976	Morphology: other	Dev and morph	Z-C2H2	Multiple developmental alterations	432	(219-323)
433	G2057	Morphology: other	Dev and morph	TEO	Multiple developmental alterations	434	(TBD)
435	G2107	Morphology: other	Dev and morph	AP2	Multiple developmental alterations	436	(TBD)

Table 4

437	G211	Morphology: other	Dev and morph	MYB-(R1)R2R3	Multiple developmental alterations	438	(24-137)
439	G2133	Morphology: other; flowering time; seed protein content	Dev and morph; flowering time	AP2	Multiple developmental alterations; late flowering; altered seed protein content	440	(11-83)
441	G2134	Morphology: other	Dev and morph	AP2	Multiple developmental alterations	442	(TBD)
443	G2151	Morphology: other; seed oil and protein content	Dev and morph; seed biochemistry	AT-hook	Multiple developmental alterations; altered seed oil and protein content	444	(93-113, 124-144)
445	G2154	Morphology: other	Dev and morph	AT-hook	Multiple developmental alterations	446	(97-119)
447	G2157	Morphology: other	Dev and morph	AT-hook	Multiple developmental alterations	448	(82-102, 164-107)
449	G2181	Morphology: other	Dev and morph	NAC	Multiple developmental alterations	450	(22-169)
451	G221	Morphology: other	Dev and morph	MYB-(R1)R2R3	Multiple developmental alterations	452	(21-125)
453	G2290	Morphology: other	Dev and morph	WRKY	Multiple developmental alterations	454	(147-205)
455	G2299	Morphology: other	Dev and morph	AP2	Multiple developmental alterations	456	(48-115)
457	G2340	Morphology: other; seed oil and protein content	Dev and morph; seed biochemistry	MYB-(R1)R2R3	Tissue necrosis; multiple developmental alterations; altered seed oil and protein content	458	(14-120)
459	G2346	Morphology: other	Dev and morph	SBP	Enlarged seedlings	460	(59-135)
461	G237	Morphology: other	Dev and morph	MYB-(R1)R2R3	Multiple developmental alterations	462	(11-113)
463	G2373	Morphology: other; seed protein content	Dev and morph; seed biochemistry	TH	Multiple developmental alterations; altered seed protein content	464	(290-350)
465	G2376	Morphology: other; seed oil protein	Dev and morph; seed biochemistry	TH	Seedling lethality; altered seed protein content	466	(79-178, 336-408)
467	G24	Morphology: other	Dev and morph	AP2	Reduced size and necrotic patches	468	(25-93)

Table 4

469	G2424	Morphology: other	Dev and morph	MYB-(R1)R2R3	Multiple developmental alterations	470	(107-219)
471	G2505	Morphology: other	Dev and morph	NAC	Lethal when overexpressed	472	(10-159)
473	G2512	Morphology: other	Dev and morph	AP2	Multiple developmental alterations	474	(79-139)
475	G2513	Morphology: other	Dev and morph	AP2	Multiple developmental alterations	476	(TBD)
477	G2519	Morphology: other	Dev and morph	HLH/MYC	Multiple developmental alterations	478	(1-65)
479	G2520	Morphology: other; seed oil and protein content	Dev and morph; seed biochemistry	HLH/MYC	Multiple developmental alterations; altered seed oil and protein content	480	(135-206)
481	G2533	Morphology: other; seed protein content	Dev and morph; seed biochemistry	NAC	Multiple developmental alterations; altered seed protein content	482	(11-186)
483	G2534	Morphology: other	Dev and morph	NAC	Lethal when overexpressed	484	(10-157)
485	G2573	Morphology: other; seed oil and protein content	Dev and morph; seed biochemistry	AP2	Multiple developmental alterations; altered seed oil and protein content	486	(31-98)
487	G2589	Morphology: other	Dev and morph	MADS	Multiple developmental alterations	488	(2-57)
489	G2687	Morphology: other	Dev and morph	AP2	Multiple developmental alterations	490	(51-120)
491	G27	Morphology: other	Dev and morph	AP2	Abnormal development, small	492	(37-104)
493	G2720	Morphology: other; seed oil and protein content	Dev and morph; seed biochemistry	MYB-(R1)R2R3	Multiple developmental alterations; altered seed oil and protein content	494	(10-114)
495	G2787	Morphology: other; seed oil content	Dev and morph; seed biochemistry	AT-hook	Multiple developmental alterations; altered seed oil content	496	(172-192, 226-247, 256-276, 290-311, 245- 366)
497	G2789	Morphology: other	Dev and morph	AT-hook	Multiple developmental alterations	498	(53-73, 121-165)
499	G31	Morphology: other	Dev and morph	AP2	Multiple developmental alterations	500	(TBD)

Table 4

501	G333	Morphology: other	Dev and morph	AP2	Multiple developmental defects	502	(50-117)
503	G342	Morphology: other; seed oil and protein content	Dev and morph; seed biochemistry	GATA/Zn	Multiple developmental alterations; altered seed oil and protein content	504	(155-190)
505	G352	Morphology: other	Dev and morph	Z-C2H2	Multiple developmental alterations	506	(99-119, 166-186)
507	G357	Morphology: other	Dev and morph	Z-C2H2	Developmental defect	508	(7-29)
509	G358	Morphology: other	Dev and morph	Z-C2H2	Lethal when overexpressed	510	(124-135, 188-210)
511	G360	Morphology: other	Dev and morph	Z-C2H2	Multiple developmental alterations	512	(42-62)
513	G362	Size; Morphology: other; trichome; flowering time; seed protein content	Dev and morph; flowering time; seed biochemistry	Z-C2H2	Reduced size; increased pigmentation in seed, embryos and other organs; ectopic trichome formation; increased trichome number; late flowering; altered protein content	514	(62-82)
515	G364	Morphology: other	Dev and morph	Z-C2H2	Developmental defect	516	(54-76)
517	G365	Morphology: other	Dev and morph	Z-C2H2	Multiple developmental alterations	518	(70-90)
519	G367	Morphology: other	Dev and morph	Z-C2H2	Lethal when overexpressed	520	(63-84)
521	G373	Morphology: other	Dev and morph	RING/C3HC4	Multiple developmental alterations	522	(129-168)
523	G396	Morphology: other; size	Dev and morph	HB	Altered leaf coloration and shape, reduced fertility; small plant	524	(159-220)
525	G431	Morphology: other	Dev and morph	HB	Developmental defect, sterile	526	(286-335)
527	G479	Morphology: other	Dev and morph	SBP	Multiple developmental alterations	528	(70-149)
529	G546	Morphology: other	Dev and morph	RING/C3H2C3	Slow growth and development; increased anthocyanin pigmentation	530	(114-155)
531	G551	Morphology: other	Dev and morph	HB	Multiple developmental alterations	532	(73-133)
533	G578	Morphology: other	Dev and morph	bZIP	Lethal when overexpressed	534	(36-96)

Table 4

535	G596	Morphology: other	Dev and morph	AT-hook	Multiple developmental alterations	536	(89-96)
537	G617	Morphology: other	Dev and morph	TEO	Multiple developmental alterations	538	(64-118)
539	G620	Morphology: other; seed protein content	Dev and morph; seed biochemistry	CAAT	Multiple developmental alterations; altered seed protein content	540	(20-118)
541	G625	Morphology: other	Dev and morph	AP2	Lethal when overexpressed	542	(52-119)
543	G658	Morphology: other	Dev and morph	MYB-(R1)R2R3	Developmental defect	544	(2-105)
545	G716	Morphology: other	Dev and morph	ARF	Multiple developmental defects	546	(24-355)
547	G725	Morphology: other	Dev and morph	GARP	Developmental defect	548	(39-87)
549	G727	Morphology: other	Dev and morph	GARP	Multiple morphological alterations	550	(226-269)
551	G740	Morphology: other	Dev and morph	Z-CLDSH	Slow growth	552	(24-42, 232-258)
553	G770	Morphology: other	Dev and morph	NAC	Multiple developmental alterations	554	(19-162)
555	G858	Morphology: other	Dev and morph	MADS	Multiple developmental alterations	556	(2-57)
557	G865	Morphology: other; seed protein content	Dev and morph; seed biochemistry	AP2	Altered morphology; increased seed protein	558	(36-103)
559	G872	Morphology: other	Dev and morph	AP2	Multiple developmental alterations	560	(18-85)
561	G904	Morphology: other	Dev and morph	RING/C3H2C3	Multiple developmental alterations	562	(117-158)
563	G910	Morphology: other; flowering time	Dev and morph; flowering time	Z-CO-like	Multiple developmental alterations; late flowering	564	(14-37, 77-103)
565	G912	Morphology: other; size; sugar sensing; flowering time	Dev and morph; sugar sensing; flowering time	AP2	Dark green color; small plant; reduced cotyledon expansion in glucose; late flowering	566	(51-118)
567	G920	Morphology: other	Dev and morph	WRKY	Multiple developmental alterations	568	(152-211)

Table 4

569	G939	Morphology: other; size	Dev and morph biochemistry	EIL	Pale seedlings on agar; reduced size	570	(97-106)
571	G963	Morphology: other; seed protein content	Dev and morph; seed biochemistry	NAC	Slowed growth rate; altered seed protein content	572	(TBD)
573	G979	Morphology: other; seed	Dev and morph	AP2	Several developmental defects; altered seed development, ripening and germination	574	(63-139, 165-233)
575	G987	Morphology: other	Dev and morph	SCR	Developmental defects	576	(428-432, 704-708)
577	G993	Morphology: other; seed protein content	Dev and morph; seed biochemistry	AP2	Multiple developmental alterations; altered seed protein content	578	(69-134)
579	G681	Morphology: other; leaf glucosinolates	Dev and morph; leaf biochemistry	MYB-(R1)R2R3	Multiple developmental alterations; overexpression results in an increase in M39480	580	(14-120)
581	G1482	Root	Dev and morph	Z-CO-like	Increased root growth	582	(5-63)
583	G225	Root; trichome	Dev and morph	MYB-related	Increased root hairs; glabrous, lack of trichomes	584	(39-76)
585	G226	Root; trichome; seed protein content	Dev and morph; seed biochemistry	MYB-related	Increased root hairs; glabrous, lack of trichomes; increased seed protein	586	(28-78)
587	G9	Root	Dev and morph	AP2	Increased root mass	588	(62-127)
589	G1040	Seed	Dev and morph	GARP	Smaller and more rounded seeds	590	(109-158)
591	G2114	Seed	Dev and morph	AP2	Increased seed size	592	(221-297, 323-393)
593	G450	Seed; size; seed protein content	Dev and morph; seed biochemistry	IAA	Increased seed size; reduced plant size; altered seed protein content	594	(TBD)
595	G584	Seed	Dev and morph	HLH/MYC	Large seeds	596	(401-494)
597	G668	Seed	Dev and morph	MYB-(R1)R2R3	Reduced seed color	598	(13-113)
599	G1050	Senescence	Dev and morph	bZIP	Delayed senescence	600	(372-425)
601	G1463	Senescence	Dev and morph	NAC	Premature senescence	602	(9-156)
603	G1944	Senescence; size; seed protein content	Dev and morph; seed biochemistry	AT-hook	Early senescence; reduced size; altered seed protein content	604	(87-100)
605	G2383	Senescence; seed protein content	Dev and morph; seed biochemistry	TEO	Early senescence; altered seed protein content	606	(89-149)
607	G571	Senescence; flowering time	Dev and morph; flowering time	bZIP	Delayed senescence; late flowering	608	(160-220)

Table 4

609	G636	Senescence; size	Dev and morph	TH	Premature senescence; reduced size	610	(55-145, 405-498)
611	G878	Senescence; flowering time	Dev and morph; flowering time	WRKY			
613	G1134	Silique	Dev and morph	HLH/MYC	Delayed senescence; late flowering	612	(250-305, 415-475)
615	G1008	Size	Dev and morph	AP2	Siliques with altered shape	614	(198-247)
617	G1020	Size	Dev and morph	AP2	Small plant	616	(96-163)
619	G1023	Size	Dev and morph	AP2	Very small T1 plants	618	(28-95)
621	G1053	Size	Dev and morph	bZIP	Reduced size	620	(128-195)
623	G1137	Size	Dev and morph	HLH/MYC	Small plant	622	(74-120)
625	G1181	Size	Dev and morph	HLH/MYC	Small T1 plants	624	(264-314)
627	G1228	Size	Dev and morph	HS	Small T1 plants	626	(24-114)
629	G1277	Size	Dev and morph	HLH/MYC	Reduced size	628	(179-233)
631	G1309	Size	Dev and morph	AP2	Small plant	630	(18-85)
			Dev and morph	MYB-(R1)R2R3	Small plant	632	(9-114)
633	G1314	Size; sugar sensing; seed protein content	Dev and morph; sugar sensing; seed biochemistry		Reduced size; reduced seedling vigor on high glucose; altered seed protein content	634	(14-116)
635	G1317	Size	Dev and morph	MYB-(R1)R2R3	Reduced size	636	(13-118)
637	G1323	Size; seed oil and protein content	Dev and morph; seed biochemistry	MYB-(R1)R2R3	Small T1 plants, dark green; decreased seed oil, increased seed protein	638	(15-116)
639	G1332	Size; trichome; seed oil and protein content	Dev and morph; seed biochemistry	MYB-(R1)R2R3	Reduced size; reduced trichome density; altered seed oil and protein content	640	(13-116)
641	G1334	Size	Dev and morph	CAAT	Small, dark green	642	(18-190)
643	G1381	Size	Dev and morph	AP2	Reduced size	644	(68-135)
645	G1382	Size	Dev and morph	WRKY	Small plant	646	(210-266, 385-437)
647	G1435	Size; flowering time	Dev and morph; flowering time				
649	G1537	Size	Dev and morph	GARP	Increased plant size; late flowering	648	(146-194)
651	G1545	Size	Dev and morph	HB	Small T1 plants with altered development	650	(14-74)
		Size; seed oil and protein content	Dev and morph	HB	Reduced size	652	(54-117)
653	G1641	Size; seed protein content	Dev and morph; seed biochemistry	MYB-related	Small plant; altered seed oil and protein content	654	(139-200)
655	G165	Size; seed protein content	Dev and morph; seed biochemistry	MADS	Reduced size; altered seed protein content	656	(7-62)

Table 4

657	G1652	Size; seed oil and protein content	Dev and morph; seed biochemistry	HLH/MYC	Reduced size; altered seed oil and protein content	658	(143-215)
659	G1655	Size	Dev and morph; biochemistry	HLH/MYC	Small plant	660	(134-192)
661	G1671	Size	Dev and morph; biochemistry	NAC	Reduced size	662	(TBD)
663	G1756	Size; seed protein content	Dev and morph; seed biochemistry	WRKY	Reduced size; altered seed protein content	664	(TBD)
665	G1757	Size; seed protein content	Dev and morph; seed biochemistry	WRKY	Small plant; altered seed protein content	666	(158-218)
667	G1782	Size	Dev and morph; biochemistry	CAAT	Small, spindly plant	668	(166-238)
669	G184	Size	Dev and morph; biochemistry	WRKY	Small plant	670	(295-352)
671	G1845	Size	Dev and morph; biochemistry	AP2	Small plant	672	(140-207)
673	G1879	Size; seed oil and protein content	Dev and morph; seed biochemistry	HLH/MYC	Reduced size; altered seed oil and protein content	674	(107-176)
675	G1888	Size	Dev and morph; biochemistry	Z-CO-like	Reduced size, dark green leaves	676	(5-50)
677	G189	Size; seed protein content	Dev and morph; seed biochemistry	WRKY	Increased leaf size; altered seed protein content	678	(240-297)
679	G1939	Size	Dev and morph; biochemistry	PCF	Reduced size	680	(40-102)
681	G194	Size	Dev and morph; biochemistry	WRKY	Small plant	682	(174-230)
683	G1943	Size	Dev and morph; biochemistry	HLH/MYC	Reduced size	684	(335-406)
685	G21	Size; seed oil and protein content	Dev and morph; seed biochemistry	AP2	Reduced size; altered seed oil and protein content	686	(97-164)
687	G2132	Size; seed oil and protein content	Dev and morph; seed biochemistry	AP2	Reduced size; altered seed oil and protein content	688	(TBD)
689	G2145	Size	Dev and morph; biochemistry	HLH/MYC	Reduced size	690	(166-243)
691	G23	Size	Dev and morph; biochemistry	AP2	Small T1 plants	692	(61-117)
693	G2313	Size	Dev and morph; biochemistry	MYB-related	Reduced size	694	(TBD)
695	G2344	Size	Dev and morph; biochemistry	CAAT	Reduced size, slow growth	696	(TBD)
697	G2430	Size	Dev and morph; biochemistry	GARP	Increased leaf size, faster development	698	(425-478)
699	G2517	Size	Dev and morph; biochemistry	WRKY	Reduced size	700	(118-234)
701	G2521	Size	Dev and morph; biochemistry	HLH/MYC	Reduced size	702	(145-213)
703	G258	Size; seed oil and protein content	Dev and morph; seed biochemistry	MYB-(R1)R2R3	Reduced size; altered seed oil and protein content	704	(24-124)

Table 4

705	G280	Size; seed protein content	Dev and morph; seed biochemistry	AT-hook	Reduced size; altered seed protein content	706	(97-104, 130-137-155-162, 185-192)
707	G3	Size	Dev and morph	AP2	Small plant	708	(28-95)
709	G343	Size	Dev and morph	GATA/Zn	Small plant	710	(178-214)
711	G363	Size	Dev and morph	Z-C2H2	Small plant	712	(87-108)
713	G370	Size	Dev and morph	Z-C2H2	Reduced size, shiny leaves	714	(97-117)
715	G385	Size	Dev and morph	HB	Small plant, short inflorescence stems, dark green	716	(60-123)
717	G439	Size	Dev and morph	AP2	Small plant	718	(110-177)
719	G440	Size	Dev and morph	AP2	Small plant	720	(122-189)
721	G5	Size	Dev and morph	AP2	Small plant	722	(149-216)
723	G550	Size	Dev and morph	Z-Dof	Small plant	724	(134-180)
725	G670	Size	Dev and morph	MYB-(R1)R2R3	Small plant	726	(14-122)
727	G760	Size	Dev and morph	NAC	Reduced size	728	(12-156)
729	G831	Size	Dev and morph	AKR	Reduced size	730	(470-591)
731	G864	Size	Dev and morph	AP2	Small plant	732	(119-186)
733	G884	Size	Dev and morph	WRKY	Reduced size	734	(227-285, 407-465)
735	G898	Size; seed oil and protein content	Dev and morph; seed biochemistry	RING/C3HC4	Reduced size; altered seed oil and protein content	736	(148-185)
737	G900	Size	Dev and morph	Z-CO-like	Reduced size	738	(6-28, 48-74)
739	G913	Size; flowering time	Dev and morph; flowering time	AP2	Small plant; late flowering	740	(62-128)
741	G937	Size	Dev and morph	GARP	Slightly reduced size	742	(197-246)
743	G960	Size	Dev and morph	NAC	Small plant	744	(13-156)
745	G991	Size; seed oil and protein content	Dev and morph; seed biochemistry	IAA	Slightly reduced size; altered seed oil and protein content	746	(7-14, 48-59, 82-115, 128-164)
747	G748	Stem; flowering time	Dev and morph; flowering time	Z-Dof	More vascular bundles in stem; late flowering	748	(112-140)
749	G247	Trichome; seed protein content	Dev and morph; seed biochemistry	MYB-(R1)R2R3	Altered trichome distribution; altered seed protein content	750	(15-116)
751	G585	Trichome	Dev and morph	HLH/MYC	Reduced trichome density	752	(436-501)
753	G634	Trichome; seed protein content	Dev and morph; seed biochemistry	TH	Increased trichome density and size; altered seed protein content	754	(62-147, 189-245)
755	G676	Trichome	Dev and morph	MYB-(R1)R2R3	Reduced trichomes	756	(17-119)

Table 4

757	G682	Trichome	Dev and morph	MYB-related	Glabrous, lack of trichomes	758	(27-63)
759	G635	Variegation	Dev and morph	TH		760	(239-323)
761	G1068	Sugar sensing	Sugar sensing	AT-hook	Reduced cotyledon expansion in glucose	762	(143-150)
763	G1225	Sugar sensing; seed oil and protein content	Sugar sensing; seed biochemistry		Better germination on sucrose and glucose media; altered seed oil and protein content	764	(78-147)
765	G1337	Sugar sensing	Sugar sensing	HLH/MYC	Decreased germination on sucrose medium	766	(9-75)
767	G1759	Sugar sensing; flowering time	Sugar sensing; flowering time	Z-CO-like	Reduced germination on high glucose	768	(2-57)
769	G1804			MADS			
771	G207	Sugar sensing	Sugar sensing	bZIP	Altered sugar sensing; late flowering	770	(357-407)
				MYB-(R1)R2R3	Decreased germination on glucose medium	772	(6-106)
773	G218	Sugar sensing; seed oil content	Sugar sensing; seed biochemistry		Reduced cotyledon expansion in glucose; altered seed oil content	774	(TBD)
				MYB-(R1)R2R3			
775	G241	Sugar sensing; seed oil and protein content	Sugar sensing; seed biochemistry		Decreased germination and growth on glucose medium; decreased seed oil, altered protein content	776	(14-114)
777	G254	Sugar sensing	Sugar sensing	MYB-(R1)R2R3	Decreased germination and growth on glucose medium	778	(62-106)
779	G26	Sugar sensing	Sugar sensing	AP2	Decreased germination and growth on glucose medium	780	(67-134)
781	G263	Sugar sensing	Sugar sensing		Decreased root growth on sucrose medium, root specific expression	782	(TBD)
783	G308	Sugar sensing	Sugar sensing	HS		784	(270-274)
785	G38	Sugar sensing	Sugar sensing	SCR	No germination on glucose medium	786	(76-143)
787	G43	Sugar sensing	Sugar sensing	AP2	Reduced germination on glucose medium	788	(104-172)
789	G536	Sugar sensing	Sugar sensing	AP2	Decreased germination and growth on glucose medium	790	(226-233)
				GF14	Decreased germination and growth on glucose medium		
791	G567	Sugar sensing; seed oil and protein content	Sugar sensing; seed biochemistry		Decreased seedling vigor on high glucose; altered seed oil and protein content	792	(210-270)
				bZIP			
793	G680	Sugar sensing; flowering time	Sugar sensing; flowering time	MYB-related	Reduced germination on glucose medium; late flowering	794	(24-70)
795	G667	Sugar sensing	Sugar sensing	AP2	Better seedling vigor on sucrose medium	796	(59-124)
797	G956	Sugar sensing	Sugar sensing	NAC	Reduced germination on glucose medium	798	(TBD)
799	G996	Sugar sensing	Sugar sensing	MYB-(R1)R2R3	Reduced germination on glucose medium	800	(14-114)
		Seed glucosinolates, oil, protein content	Seed biochemistry				
801	G1946				Increase in M3950; increased oil content; decreased protein content	802	(32-130)
803	G217	Seed oil composition	Seed biochemistry	HS			
				MYB-related	Increase in 20:2	804	(8-67)

Table 4

805	G2192	Seed oil composition	Seed biochemistry	bZIP-NIN	Altered composition	806	(600-700)
807	G504	Seed oil composition;	Seed biochemistry	NAC	Altered seed oil composition and content; altered seed protein content	808	(TBD)
809	G622	Seed oil composition	Seed biochemistry	ABI3/VP-1	Decreased 18:2 fatty acid	810	(TBD)
811	G778	Seed oil composition	Seed biochemistry	HLH/MYC	Increased seed 18:1 fatty acid	812	(220-267)
813	G791	Seed oil composition	Seed biochemistry	HLH/MYC	Altered seed fatty acid composition	814	(75-143)
815	G861	Seed oil composition; seed oil content	Seed biochemistry	MADS	Increase in 16:1; altered seed oil content	816	(2-57)
817	G938	Seed oil composition	Seed biochemistry	EIL	Altered seed fatty acid composition	818	(96-104)
819	G965	Seed oil composition	Seed biochemistry	HB	Increase in 18:1	820	(423-486)
821	G1143	Seed oil and protein content	Seed biochemistry	HLH/MYC	Altered seed oil and protein content	822	(33-82)
823	G1190	Seed oil content	Seed biochemistry	AKR	Increased content	824	(entire protein)
825	G1198	Seed oil and protein content	Seed biochemistry	bZIP	Altered seed oil and protein content	826	(173-223)
827	G1226	Seed oil and protein content	Seed biochemistry	HLH/MYC	Altered seed oil and protein content	828	(115-174)
829	G1451	Seed oil content	Seed biochemistry	ARF	Altered seed oil content	830	(22-357)
831	G1478	Seed oil and protein content; flowering time	Seed biochemistry; flowering time	Z-CO-like	Altered seed oil, protein content; late flowering	832	(32-76)
833	G1496	Seed oil content	Seed biochemistry	HLH/MYC	Altered seed oil content	834	(184-248)
835	G1526	Seed oil content	Seed biochemistry	SWI/SNF	Increased seed oil content	836	(493-620, 864-1006)
837	G1543	Seed oil content	Seed biochemistry	HB	Decreased seed oil	838	(135-195)
839	G162	Seed oil and protein content	Seed biochemistry	MADS	Altered seed oil content; altered seed oil and protein content	840	(2-57)

Table 4

841	G1640	Seed oil content	Seed biochemistry	MYB-(R1)R2R3	Increased seed oil	842	(14-115)
843	G1644	Seed oil and protein content	Seed biochemistry	MYB-(R1)R2R3	Altered seed oil, protein content	844	(39-102)
845	G1646	Seed oil content	Seed biochemistry	CAAT	Altered seed oil content	846	(72-162)
847	G1672	Seed oil content	Seed biochemistry	NAC	Altered seed oil content	848	(41-194)
849	G1677	Seed oil and protein content	Seed biochemistry	NAC	Altered seed oil, protein content	850	(17-181)
851	G1765	Seed oil and protein content	Seed biochemistry	NAC	Altered seed oil and protein content	852	(20-140)
853	G1777	Seed oil and protein content	Seed biochemistry	RING/C3HC4	Increased oil, decreased protein content	854	(124-247)
855	G1793	Seed oil content	Seed biochemistry	AP2	Increased seed oil content	856	(179-255, 281-349)
857	G180	Seed oil content	Seed biochemistry	WRKY	Decreased seed oil content	858	(118-174)
859	G192	Seed oil and protein content; flowering time	Seed biochemistry; flowering time	WRKY	Altered seed oil and protein content; late flowering	860	(128-185)
861	G1948	Seed oil and protein content	Seed biochemistry	AKR	Altered seed oil and protein content	862	(entire protein)
863	G2123	Seed oil and protein content	Seed biochemistry	GF14	Altered seed oil and protein content	864	(99-109)
865	G2138	Seed oil content	Seed biochemistry	AP2	Increased seed oil content	866	(TBD)
867	G2139	Seed oil content	Seed biochemistry	MADS	Increased seed content	868	(14-69)
869	G2343	Seed oil content	Seed biochemistry	MYB-(R1)R2R3	Altered seed oil content	870	(14-116)
871	G265	Seed oil and protein content	Seed biochemistry	HS	Altered seed oil and protein content	872	(11-105)
873	G2792	Seed oil content	Seed biochemistry	HLH/MYC	Increased seed oil content	874	(190-258)
875	G2830	Seed oil and protein content	Seed biochemistry	Z-C2H2	Altered seed oil and protein content	876	(245-266)
877	G286	Seed oil and protein content	Seed biochemistry	ENBP	Altered seed oil and protein content	878	(TBD)

Table 4

879	G291	Seed oil content	Seed biochemistry	MISC	Increased seed oil content	880	(132-160)
881	G427	Seed oil and protein content	Seed biochemistry	HB	Increased oil content; decreased protein content	882	(307-370)
883	G509	Seed oil and protein content	Seed biochemistry	NAC	Altered seed oil and protein content	884	(13-169)
885	G519	Seed oil and protein content	Seed biochemistry	NAC	Altered seed oil and protein content	886	(11-104)
887	G561	Seed oil content	Seed biochemistry	bZIP	Altered seed oil content	888	(248-308)
889	G590	Seed oil and protein content	Seed biochemistry	HLH/MYC	Altered seed oil and protein content	890	(202-254)
891	G818	Seed oil content	Seed biochemistry	HS	Increased content	892	(70-162)
893	G849	Seed oil and protein content	Seed biochemistry	BPF-1	Increased seed oil, altered protein content	894	(324-413, 504-583)
895	G892	Seed oil and protein content	Seed biochemistry	RING/C3H2C3	Altered seed oil, protein content	896	(177-270)
897	G961	Seed oil content	Seed biochemistry	NAC	Altered seed oil content	898	(15-140)
899	G1465	Seed oil and protein content	Seed biochemistry	NAC	Altered seed oil and protein content	900	(242-306)
901	G425	Seed oil content	Seed biochemistry	HB	Altered seed oil content	902	(TBD)
903	G347	Seed oil and protein content	Seed biochemistry	Z-LSDlike	Altered seed oil and protein content	904	(9-39, 50-70, 80-127)
905	G1512	Seed oil and protein content	Seed biochemistry	RING/C3HC4	Altered seed oil and protein content	906	(39-93)
907	G2069	Seed oil and protein content	Seed biochemistry	bZIP	Altered seed oil and protein content	908	(TBD)
909	G1852	Seed oil content	Seed biochemistry	AKR	Altered seed oil content	910	(1-601)
911	G1793	Seed oil content	Seed biochemistry	AP2	Altered seed oil content	912	(179-255, 281-349)
913	G761	Seed oil and protein content	Seed biochemistry	NAC	Altered seed oil and protein content	914	(10-156)
915	G1056	Seed oil content	Seed biochemistry	bZIP	Altered seed oil content	916	(183-246)
917	G1447	Seed oil content	Seed biochemistry	MISC	Altered seed oil content	918	(3-54, 124-156)

Table 4

919	G323	Seed oil and protein content	Seed biochemistry	RING/C3HC4	Altered seed oil and protein content	920	(48-96)
921	G176	Seed oil content	Seed biochemistry	WRKY	Altered seed oil content	922	(117-173, 234-290)
923	G174	Seed oil and protein content	Seed biochemistry	WRKY	Altered seed oil and protein content	924	(111-166, 283-339)
925	G715	Seed oil content	Seed biochemistry	CAAT	Altered seed oil content	926	(60-132)
927	G588	Seed oil and protein content	Seed biochemistry	HLH/MYC	Altered seed oil and protein content	928	(309-376)
929	G1758	Seed oil and protein content	Seed biochemistry	WRKY	Altered seed oil and protein content	930	(109-165)
931	G2148	Seed oil content	Seed biochemistry	HLH/MYC	Altered seed oil content	932	(130-268)
933	G2379	Seed oil content	Seed biochemistry	TH	Altered seed oil content	934	(19-110, 173-232)
935	G1462	Seed oil content	Seed biochemistry	NAC	Altered seed oil content	936	(TBD)
937	G1211	Seed oil and protein content	Seed biochemistry	MISC	Altered seed oil and protein content	938	(123-179)
939	G1048	Seed oil content	Seed biochemistry	bZIP	Altered seed oil content	940	(138-190)
941	G986	Seed oil content	Seed biochemistry	WRKY	Altered seed oil content	942	(146-203)
943	G789	Seed oil content	Seed biochemistry	HLH/MYC	Altered seed oil content	944	(253-313)
945	G2085	Seed oil and protein content	Seed biochemistry	RING/C3HC4	Altered seed oil and protein content	946	(TBD)
947	G1783	Seed oil and protein content	Seed biochemistry	MYB-related	Altered seed oil and protein content	948	(81-129)
949	G2072	Seed oil and protein content	Seed biochemistry	bZIP	Altered seed oil and protein content	950	(90-149)
951	G931	Seed oil and protein content	Seed biochemistry	CAAT	Altered seed oil and protein content	952	(TBD)
953	G278	Seed oil and protein content	Seed biochemistry	AKR	Altered seed oil and protein content	954	(2-593)
955	G2421	Seed oil content	Seed biochemistry	MYB-(R1)R2R3	Altered seed oil content	956	(9-110)
957	G2032	Seed oil content	Seed biochemistry	AKR	Altered seed oil content	958	(entire protein)

Table 4

959	G1396	Seed oil and protein content	Seed biochemistry	S1FA	Altered seed oil and protein content	960	(TBD)
961	G619	Seed oil and protein content	Seed biochemistry	ARF	Altered seed oil and protein content	962	(64-406)
963	G2295	Seed oil content	Seed biochemistry	MADS	Altered seed oil content	964	(2-57)
965	G312	Seed oil content	Seed biochemistry	SCR	Altered seed oil content	966	(320-336)
967	G1444	Seed oil and protein content	Seed biochemistry	GRF-like	Altered seed oil and protein content	968	(168-193)
969	G801	Seed oil content	Seed biochemistry	PCF	Altered seed oil content	970	(32-93)
971	G1950	Seed oil content	Seed biochemistry	AKR	Altered seed oil content	972	(65-228)
973	G958	Seed oil and protein content	Seed biochemistry	NAC	Altered seed oil and protein content	974	(7-156)
975	G1037	Seed oil and protein content	Seed biochemistry	GARP	Altered seed oil and protein content	976	(11-134, 200-248)
977	G2065	Seed oil content	Seed biochemistry	MADS	Altered seed oil content	978	(TBD)
979	G2137	Seed oil and protein content	Seed biochemistry	WRKY	Altered seed oil and protein content	980	(109-168)
981	G746	Seed oil content	Seed biochemistry	RING/C3HC4	Altered seed oil content	982	(139-178)
983	G2701	Seed oil and protein content	Seed biochemistry	MYB-related	Altered seed oil and protein content	984	(33-81, 129-183)
985	G1819	Seed oil content	Seed biochemistry	CAAT	Altered seed oil content	986	(46-188)
987	G1227	Seed oil and protein content	Seed biochemistry	HLH/MYC	Altered seed oil and protein content	988	(183-244)
989	G2417	Seed oil content	Seed biochemistry	GARP	Altered seed oil content	990	(235-285)
991	G2116	Seed oil content	Seed biochemistry	bZIP	Altered seed oil content	992	(150-210)
993	G647	Seed oil content	Seed biochemistry	Z-C3H	Altered seed oil content	994	(77-192)
995	G974	Seed oil and protein content	Seed biochemistry	AP2	Altered seed oil and protein content	996	(81-140)
997	G1419	Seed protein content	Seed biochemistry	AP2	Increased seed protein	998	(69-137)

Table 4

999	G1634	Seed protein content	Seed biochemistry	MYB-related	Altered seed protein content	1000	(129-180)
1001	G1637	Seed protein content	Seed biochemistry	MYB-related	Altered seed protein content	1002	(109-173)
1003	G1818	Seed protein content; flowering time	Seed biochemistry; flowering time	CAAT	Increased protein content; late flowering	1004	(36-113)
1005	G1820	Seed oil and protein content	Seed biochemistry	CAAT	Altered seed oil, protein content	1006	(70-133)
1007	G1903	Seed oil and protein content	Seed biochemistry	Z-Dof	Altered seed oil and protein content	1008	(134-180)
1009	G371	Seed oil and protein content	Seed biochemistry	RING/C3HC4	Altered seed oil and protein content	1010	(21-74)
1011	G597	Seed protein content	Seed biochemistry	AT-hook	Altered seed protein content	1012	(97-104, 137-144)
1013	G1009	Seed protein content	Seed biochemistry	AP2	Altered seed protein content	1014	(201-277, 303-371)
1015	G170	Seed protein content	Seed biochemistry	MADS	Altered seed protein content	1016	(2-57)
1017	G1768	Seed protein content	Seed biochemistry	SCR	Altered seed protein content	1018	(54-413)
1019	G185	Seed protein content	Seed biochemistry	WRKY	Altered seed protein content	1020	(113-172)
1021	G1931	Seed protein content	Seed biochemistry	WRKY	Altered seed protein content	1022	(114-170)
1023	G2543	Seed protein content	Seed biochemistry	HB	Altered seed protein content	1024	(31-91)
1025	G264	Seed protein content	Seed biochemistry	HS	Altered seed protein content	1026	(24-114)
1027	G32	Seed protein content	Seed biochemistry	AP2	Altered seed protein content	1028	(17-84)
1029	G436	Seed protein content	Seed biochemistry	HB	Altered seed protein content	1030	(22-85)
1031	G556	Seed protein content	Seed biochemistry	bZIP	Altered seed protein content	1032	(83-143)
1033	G1420	Seed protein content	Seed biochemistry	WRKY	Altered seed protein content	1034	(221-280)
1035	G1412	Seed protein content	Seed biochemistry	NAC	Altered seed protein content	1036	(17-159)

Table 4

1037	G738	Seed protein content	Seed biochemistry	Z-Dof	Altered seed protein content	1038	(351-393)
1039	G2426	Seed protein content	Seed biochemistry	MYB-(R1)R2R3	Altered seed protein content	1040	(14-114)
1041	G1524	Seed protein content	Seed biochemistry	RING/C3HC4	Altered seed protein content	1042	(49-110)
1043	G1243	Seed protein content	Seed biochemistry	SWI/SNF	Altered seed protein content	1044	(216-609)
1045	G631	Seed protein content	Seed biochemistry	bZIP	Altered seed protein content	1046	(TBD)
1047	G1909	Seed protein content	Seed biochemistry	Z-Dof	Altered seed protein content	1048	(23-51)
1049	G1663	Seed protein content	Seed biochemistry	PCF	Altered seed protein content	1050	(TBD)
1051	G1231	Seed protein content	Seed biochemistry	Z-C4HC3	Altered seed protein content	1052	(TBD)
1053	G227	Seed protein content	Seed biochemistry	MYB-(R1)R2R3	Altered seed protein content	1054	(13-112)
1055	G1842	Seed protein content	Seed biochemistry	MADS	Altered seed protein content	1056	(2-57)
1057	G1505	Seed protein content	Seed biochemistry	GATA/Zn	Altered seed protein content	1058	(TBD)
1059	G657	Seed protein content	Seed biochemistry	MYB-(R1)R2R3	Altered seed protein content	1060	(TBD)
1061	G1959	Seed protein content	Seed biochemistry	GARP	Altered seed protein content	1062	(46-97)
1063	G2180	Seed protein content	Seed biochemistry	NAC	Altered seed protein content	1064	(7-156)
1065	G1817	Seed protein content	Seed biochemistry	PMR	Altered seed protein content	1066	(47-331)
1067	G1649	Seed protein content	Seed biochemistry	HLH/MYC	Altered seed protein content	1068	(225-295)
1069	G2131	Seed protein content	Seed biochemistry	AP2	Altered seed protein content	1070	(50-186, 112-183)
1071	G215	Seed protein content	Seed biochemistry	MYB-related	Altered seed protein content	1072	(TBD)
1073	G1508	Seed protein content	Seed biochemistry	GATA/Zn	Altered seed protein content	1074	(38-63)
1075	G2110	Seed protein content	Seed biochemistry	WRKY	Altered seed protein content	1076	(239-298)

Table 4

1077	G2442	Seed protein content	Seed biochemistry	RING/C3HC4	Altered seed protein content	1078	(220-246)
1079	G1051	Flowering time	Flowering time	bZIP	Late flowering	1080	(189-250)
1081	G1052	Flowering time	Flowering time	bZIP	Late flowering	1082	(201-261)
1083	G1079	Flowering time;	Flowering time;	BZIP2	Late flowering; altered seed protein content	1084	(1-50)
1085	G1335	Flowering time	Flowering time	Z-CLDSH	Late flowering, slow growth	1086	(24-43, 131-144, 185-203)
1087	G157	Flowering time	Flowering time	MADS	Altered flowering; significant overexpression delays flowering time	1088	(2-57)
1089	G1895	Flowering time	Flowering time	Z-Dof	Late flowering	1090	(55-110)
1091	G1900	Flowering time	Flowering time	Z-Dof	Late flowering	1092	(54-106)
1093	G2007	Flowering time; seed protein content	Flowering time; seed				
1095	G2014	Flowering time	biochemistry	MYB-(R1)R2R3	Late flowering; altered seed protein content	1094	(TBD)
1097	G2155	Flowering time	Flowering time	MYB-related	Late flowering	1096	(22-71)
1099	G234	Flowering time	Flowering time	AT-hook	Late flowering	1098	(18-38)
1101	G361	Flowering time	Flowering time	MYB-(R1)R2R3	Late flowering, small plant	1100	(14-115)
1103	G562	Flowering time	Flowering time	Z-C2H2	Late flowering	1102	(43-63)
1105	G591	Flowering time	Flowering time	bZIP	Late flowering	1104	(253-315)
1107	G8	Flowering time	Flowering time	HLH/MYC	Late flowering	1106	(143-240)
		Flowering time; seed protein content	Flowering time; seed	AP2	Late flowering	1108	(151-217, 243-296)
1109	G859	Flowering time; seed protein content	biochemistry	MADS	Late flowering; altered seed protein content	1110	(TBD)
1111	G878	Flowering time	Flowering time	WRKY	Late flowering	1112	(250-305, 415-475)
1113	G971	Flowering time	Flowering time	AP2	Late flowering	1114	(120-186)
1115	G975	Flowering time; morphology; other	Flowering time; dev and morph	AP2	Late flowering; glossy leaves	1116	(4-71)
1117	G994	Flowering time	Flowering time	MYB-(R1)R2R3	Late flowering, small	1118	(14-123)
1119	G2347	Flowering time	Flowering time	SBP	Late flowering, small	1120	(60-136)
1121	G2010	Flowering time	Flowering time	SBP	Late flowering	1122	(53-127)

Table 5

SEQ ID NO	GID	Test Sequence ID	Smallest Sum Probability	Test Sequence Species	Test Sequence GenBank Annotation
859	G192	AW596933	7.70E-40	[Glycine max]	sJ84f07.y1 Gm-c1034 Glycine max cDNA clone GENO
859	G192	AV423663	2.40E-39	[Lotus japonicus]	AV423663 Lotus japonicus young plants (two-
859	G192	BI422074	4.50E-34	[Lycopersicon esculentum]	EST532740 tomato callus, TAMU Lycop
859	G192	AW447931	1.40E-27	[Triticum aestivum]	BRY_1082 BRY Triticum aestivum cDNA clone
859	G192	BE998060	2.60E-24	[Medicago truncatula]	EST429783 GVSJN Medicago truncatula cDNA
859	G192	AC018727	1.70E-23	[Oryza sativa]	chromosome 10 clone OSJNBa0056G17. *** SEQUENC
859	G192	BG600477	1.00E-20	[Solanum tuberosum]	EST505372 cSTS Solanum tuberosum cDNA clo
859	G192	BG356878	2.80E-16	[Sorghum bicolor]	OV2_11 B04.g1_A002 Ovary 2 (OV2) Sorghum bi
859	G192	gi12039364	1.10E-31	[Oryza sativa]	putative DNA-binding protein.
859	G192	gi4894963	3.30E-14	[Avena sativa]	DNA-binding protein WRKY3.
859	G192	gi1432056	5.80E-14	[Petroselinum crispum]	WRKY3.
859	G192	gi4760596	2.60E-13	[Nicotiana tabacum]	DNA-binding protein NtWRKY3.
859	G192	gi11993901	1.40E-12	[Dactylis glomerata]	somatic embryogenesis related protein.
859	G192	gi927025	7.60E-09	[Cucumis sativus]	SPF1-like DNA-binding protein.
859	G192	gi13620227	8.40E-09	[Lycopersicon esculentum]	hypothetical protein.
859	G192	gi3420906	2.80E-08	[Pimpinella brachycarpa]	zinc finger protein; WRKY1.
859	G192	gi1159877	4.70E-08	[Avena fatua]	DNA-binding protein.
859	G192	gi484261	1.60E-07	[Ipomoea batatas]	SPF1 protein.
801	G1946	LPHSF8	1.10E-119	[Lycopersicon peruvianum]	L.peruvianum Lp-hsf8 mRNA for heat
801	G1946	AC087771	4.10E-112	[Medicago truncatula]	clone 8D15, *** SEQUENCING IN PROGRESS
801	G1946	LEHSF8	5.90E-103	[Lycopersicon esculentum]	L.esculentum Le-hsf8 gene for heat
801	G1946	AW569138	3.10E-75	[Glycine max]	si63g09.y1 Gm-r1030 Glycine max cDNA clone GENO
801	G1946	BG890899	1.30E-70	[Solanum tuberosum]	EST516750 cSTD Solanum tuberosum cDNA clo
801	G1946	AC027658	4.60E-53	[Oryza sativa]	subsp. japonica BAC nbxb0006113, chromosome 10
801	G1946	AV833112	4.90E-52	[Hordeum vulgare subsp. vulgare]	AV833112 K. Sato unpublished
801	G1946	gi19492	2.80E-121	[Lycopersicon peruvianum]	heat shock transcription factor 8
801	G1946	gi19260	5.10E-106	[Lycopersicon esculentum]	heat stress transcription factor
801	G1946	gi662924	2.00E-47	[Glycine max]	heat shock transcription factor 21.
801	G1946	gi5821138	9.70E-46	[Nicotiana tabacum]	heat shock factor.
801	G1946	gi11761077	2.90E-40	[Oryza sativa]	putative heat shock factor protein 1 (HSF 1)
801	G1946	gi886742	3.20E-40	[Zea mays]	heat shock factor.
801	G1946	gi17158882	2.70E-38	[Medicago sativa]	heat shock transcription factor.
801	G1946	gi3550588	1.90E-30	[Pisum sativum]	heat shock transcription factor (HSFA).

Table 5

801	G1946	gi100546	0.46	[Avena sativa]	avenin precursor - oat.
801	G1946	gi14190783	1	[Apium graveolens]	putative phloem transcription factor M1.
239	G375	AW696439	3.40E-33	[Medicago truncatula]	NF106B07ST1F1060 Developing stem Medica
239	G375	BG595870	1.90E-31	[Solanum tuberosum]	EST494548 cSTS Solanum tuberosum cDNA clo
239	G375	AI899263	3.70E-31	[Lycopersicon esculentum]	EST288706 tomato ovary, TAMU Lycopen
239	G375	NTBBF3	4.00E-31	[Nicotiana tabacum]	N.tabacum mRNA for zinc finger protein, B
239	G375	BG405482	2.70E-30	[Glycine max]	sac44a11.y1 Gm-c1062 Glycine max cDNA clone GEN
239	G375	AB028130	3.30E-30	[Oryza sativa]	mRNA for Dof zinc finger protein, complete cds
239	G375	AB026297	7.30E-28	[Pisum sativum]	mRNA for elicitor-responsive Dof protein ERDP
239	G375	HVBPF	1.10E-27	[Hordeum vulgare]	mRNA for DNA binding protein BPBF.
239	G375	BG263089	1.70E-27	[Triticum aestivum]	WHE2337_A02_A03ZS Wheat pre-anthesis spik
239	G375	ZMU82230	4.20E-27	[Zea mays]	endosperm-specific prolamin box binding factor (PB
239	G375	gi4996640	1.90E-37	[Oryza sativa]	Dof zinc finger protein.
239	G375	gi3777436	8.10E-35	[Hordeum vulgare]	DNA binding protein.
239	G375	gi2393775	1.10E-33	[Zea mays]	prolamin box binding factor.
239	G375	gi1360088	2.00E-33	[Nicotiana tabacum]	Zn finger protein.
239	G375	gi3790264	4.30E-32	[Triticum aestivum]	PBF protein.
239	G375	gi6092016	1.30E-29	[Pisum sativum]	elicitor-responsive Dof protein ERDP.
239	G375	gi7688355	5.60E-29	[Solanum tuberosum]	Dof zinc finger protein.
239	G375	gi1669341	4.60E-20	[Cucurbita maxima]	AOBP (ascorbate oxidase promoter-binding
239	G375	gi3929325	5.50E-18	[Dendrobium grex Madame Thong-In]	putative DNA-binding prot
239	G375	gi19547	5.50E-06	[Medicago sativa subsp. falcata]	environmental stress and a
273	G1255	AC087181	1.60E-46	[Oryza sativa]	chromosome 3 clone OSJNBa0018H01, *** SEQUENCI
273	G1255	BG239774	4.50E-33	[Glycine max]	sab74c03.y1 Gm-c1032 Glycine max cDNA clone GEN
273	G1255	BG321336	1.70E-32	[Descurainia sophia]	Ds01_06h10_A Ds01_AAFEC ECORC_cold_stress
273	G1255	AI772841	2.90E-30	[Lycopersicon esculentum]	EST253941 tomato resistant, Cornell
273	G1255	BF480245	4.60E-29	[Mesembryanthemum crystallinum]	L0-2152T3 Ice plant Lambda Un
273	G1255	AW688119	2.10E-28	[Medicago truncatula]	NF002E07ST1F1000 Developing stem Medica
273	G1255	BF266327	1.80E-26	[Hordeum vulgare]	HV_CEA0014N02f Hordeum vulgare seedling gre
273	G1255	AW671538	5.80E-25	[Sorghum bicolor]	LG1_348_B08.b1_A002 Light Grown 1 (LG1) Sor
273	G1255	BI072021	5.30E-20	[Populus tremula x Populus tremuloides]	C067P76U Populus stra
273	G1255	BG273908	4.90E-19	[Vitis vinifera]	EST 110 Green Grape berries Lambda Zap II Lj
273	G1255	gi13702811	3.70E-52	[Oryza sativa]	putative zinc finger protein.
273	G1255	gi11037311	4.00E-21	[Brassica nigra]	constans-like protein.
273	G1255	gi2303683	1.10E-19	[Brassica napus]	unnamed protein product.
273	G1255	gi4091804	2.30E-18	[Malus x domestica]	CONSTANS-like protein 1.

Table 5

273	G1255	gi3341723	4.30E-18	[Raphanus sativus]	CONSTANS-like 1 protein.
273	G1255	gi10946337	5.20E-17	[Ipomoea nil]	CONSTANS-like protein.
273	G1255	gi4557093	3.30E-15	[Pinus radiata]	zinc finger protein.
273	G1255	gi8132543	0.97	[Chloroplast Zamia furfuracea]	cytochrome b559 alpha subunit.
273	G1255	gi11795	0.99	[Nicotiana tabacum]	put. psbE protein (aa 1-83).
273	G1255	gi65646	0.99	[Chloroplast Nicotiana tabacum]	cytochrome b559 component p
557	G865	BE419451	3.70E-32	[Triticum aestivum]	WWS012.C2R000101 ITEC WWS Wheat Scutellum
557	G865	AW560968	1.10E-28	[Medicago truncatula]	EST316016 DSIR Medicago truncatula cDNA
557	G865	AW782252	1.20E-26	[Glycine max]	sm03d11.y1 Gm-c1027 Glycine max cDNA clone GENO
557	G865	BI421895	3.60E-25	[Lycopersicon esculentum]	EST532561 tomato callus, TAMU Lycopersicon
557	G865	BE642320	1.60E-24	[Ceratopteris richardii]	Cr12_5 L17 SP6 Ceratopteris Spore LI
557	G865	BE494041	1.60E-24	[Secale cereale]	WHE1277 B09 D17ZS Secale cereale anther cDNA
557	G865	D39914	2.60E-24	[Oryza sativa]	RICS1576A Rice shoot Oryza sativa cDNA, mRNA s
557	G865	AV428124	9.00E-23	[Lotus japonicus]	AV428124 Lotus japonicus young plants (two- Tobacco mRNA for EREBP-2, complete cds.
557	G865	TOBBY4D	1.80E-21	[Nicotiana tabacum]	ERF1.
557	G865	gi1208495	2.40E-23	[Nicotiana tabacum]	ethylene-responsive element binding
557	G865	gi8809571	5.10E-23	[Nicotiana sylvestris]	Pti4.
557	G865	gi3342211	1.40E-22	[Lycopersicon esculentum]	AP2-related transcription f
557	G865	gi7528276	1.70E-22	[Mesembryanthemum crystallinum]	Putative AP2 domain containing protein.
557	G865	gi15217291	7.80E-22	[Oryza sativa]	AP2 domain containing protein.
557	G865	gi3264767	2.70E-21	[Prunus armeniaca]	AP2-domain DNA-binding protein.
557	G865	gi8980313	2.10E-20	[Catharanthus roseus]	apetala2 domain-containing protein.
557	G865	gi8571476	9.30E-20	[Atriplex hortensis]	DNA binding protein homolog.
557	G865	gi1688233	1.40E-19	[Solanum tuberosum]	ethylene-responsive element binding
557	G865	gi6478845	1.80E-19	[Matricaria chamomilla]	BOHOJ67TR BOHO Brassica oleracea genomic
23	G2509	BH577856	2.50E-29	[Brassica oleracea]	sak01e08.y1 Gm-c1074 Glycine max cDNA clone SOY
23	G2509	BM269574	5.90E-28	[Glycine max]	WWS012.C2R000101 ITEC WWS Wheat Scutellum
23	G2509	BE419451	2.20E-27	[Triticum aestivum]	EST249507 tomato ovary, TAMU Lycopersicon
23	G2509	AI483636	7.80E-27	[Lycopersicon esculentum]	EST316016 DSIR Medicago truncatula cDNA
23	G2509	AW560968	8.90E-27	[Medicago truncatula]	Cr12_5 L17 SP6 Ceratopteris Spore LI
23	G2509	BE642320	4.30E-26	[Ceratopteris richardii]	chromosome 1 clone P0677H08, *** SEQUENCING IN
23	G2509	AP003286	1.00E-25	[Oryza sativa]	WHE1277 B09 D17ZS Secale cereale anther cDNA
23	G2509	BE494041	3.20E-25	[Secale cereale]	HVSMEH0102106f Hordeum vulgare 5-45 DAP spi
23	G2509	BE602106	1.10E-24	[Hordeum vulgare]	AV428124 Lotus japonicus young plants (two- AP2 domain containing protein.
23	G2509	AV428124	1.00E-23	[Lotus japonicus]	
23	G2509	gi3264767	4.00E-27	[Prunus armeniaca]	

Table 5

23	G2509	gi12003376	1.40E-23	[Nicotiana tabacum]	Avr9/Cf-9 rapidly elicited protein 1.
23	G2509	gi14140141	2.30E-23	[Oryza sativa]	putative AP2-related transcription factor.
23	G2509	gi1688233	5.40E-23	[Solanum tuberosum]	DNA binding protein homolog.
23	G2509	gi4099921	2.60E-22	[Stylosanthes hamata]	EREBP-3 homolog.
23	G2509	gi8809571	7.80E-22	[Nicotiana sylvestris]	ethylene-responsive element binding
23	G2509	gi3342211	1.00E-21	[Lycopersicon esculentum]	Pti4.
23	G2509	gi7528276	2.70E-21	[Mesembryanthemum crystallinum]	AP2-related transcription f
23	G2509	gi17385636	1.90E-20	[Matricaria chamomilla]	ethylene-responsive element binding
23	G2509	gi18496063	3.30E-20	[Fagus sylvatica]	ethylene responsive element binding prote
1119	G2347	BI931517	5.30E-31	[Lycopersicon esculentum]	EST551406 tomato flower, 8 mm to pr
1119	G2347	BE058432	4.20E-29	[Glycine max]	sn16a06.y1 Gm-c1016 Glycine max cDNA clone GENO
1119	G2347	AMSPB1	1.80E-28	[Antirrhinum majus]	A.majus mRNA for squamosa-promoter bindin
1119	G2347	BG525285	5.70E-28	[Stevia rebaudiana]	48-3 Stevia field grown leaf cDNA Stevia
1119	G2347	L38193	4.60E-27	[Brassica rapa]	BNAF1025E Mustard flower buds Brassica rapa c
1119	G2347	BG455868	6.40E-27	[Medicago truncatula]	NF068F05PL1F1045 Phosphate starved leaf
1119	G2347	BG097153	1.70E-24	[Solanum tuberosum]	EST461672 potato leaves and petioles Sola
1119	G2347	BF482644	1.60E-23	[Triticum aestivum]	WHE2301-2304_A21_A21ZS Wheat pre-anthesis
1119	G2347	AW747167	2.30E-23	[Sorghum bicolor]	WS1_66_F11.b1_A002 Water-stressed 1 (WS1) S
1119	G2347	BG442540	2.50E-23	[Gossypium arboreum]	GA_Ea0017G06f Gossypium arboreum 7-10 d
1119	G2347	gi1183864	1.50E-31	[Antirrhinum majus]	squamosa-promoter binding protein 2.
1119	G2347	gi5931786	3.40E-25	[Zea mays]	SBP-domain protein 5.
1119	G2347	gi8468036	1.40E-21	[Oryza sativa]	Similar to Arabidopsis thaliana chromosome 2
1119	G2347	gi9087308	6.60E-09	[Mitochondrion Beta vulgaris var. altissima]	orf102a.
1119	G2347	gi7209500	0.83	[Brassica rapa]	S-locus pollen protein.
43	G988	CRU303349	3.10E-208	[Capsella rubella]	ORF1, ORF2, ORF3, ORF4, ORF5 and ORF6 (pa
43	G988	A84072	4.50E-86	[Lycopersicon esculentum]	Sequence 1 from Patent WO9846759.
43	G988	A84080	3.30E-85	[Solanum tuberosum]	Sequence 9 from Patent WO9846759.
43	G988	AP003944	1.30E-57	[Oryza sativa]	chromosome 6 clone OJ1126_F05, *** SEQUENCING
43	G988	AX081276	2.80E-43	[Brassica napus]	Sequence 1 from Patent WO0109356.
43	G988	ZMA242530	1.50E-40	[Zea mays]	partial d8 gene for gibberellin response modulator
43	G988	AX005804	2.50E-37	[Triticum aestivum]	Sequence 13 from Patent WO9909174.
43	G988	AB048713	9.10E-33	[Pisum sativum]	PsSCR mRNA for SCARECROW, complete cds.
43	G988	AW774515	2.00E-29	[Medicago truncatula]	EST333666 KV3 Medicago truncatula cDNA
43	G988	BE822458	1.20E-27	[Glycine max]	GM700017A20H12 Gm-r1070 Glycine max cDNA clone
43	G988	gi13620166	8.00E-211	[Capsella rubella]	hypothetical protein.
43	G988	gi4160441	1.40E-87	[Lycopersicon esculentum]	lateral suppressor protein.

Table 5

43	G988	gi10178637	2.20E-48	[Zea mays]	SCARECROW.
43	G988	gi6970472	1.20E-47	[Oryza sativa]	OsGAI.
43	G988	gi5640157	2.80E-45	[Triticum aestivum]	gibberellin response modulator.
43	G988	gi13170126	7.10E-45	[Brassica napus]	unnamed protein product.
43	G988	gi13365610	1.10E-40	[Pisum sativum]	SCARECROW.
43	G988	gi14318115	1.10E-14	[Zea mays subsp. mays]	gibberellin response modulator.
43	G988	gi14318165	7.30E-14	[Tripsacum dactyloides]	gibberellin response modulator.
43	G988	gi347457	2.40E-05	[Glycine max]	hydroxyproline-rich glycoprotein.
459	G2346	AMA011622	3.10E-35	[Antirrhinum majus]	mRNA for squamosa promoter binding
459	G2346	AW691786	1.80E-26	[Medicago truncatula]	NF044B06ST1F1000 Developing stem Medica
459	G2346	AQ273505	7.00E-25	[Oryza sativa]	nxb0030003f CUGI Rice BAC Library Oryza sativ
459	G2346	AW932595	7.90E-24	[Lycopersicon esculentum]	EST358438 tomato fruit mature green
459	G2346	BG593787	9.50E-24	[Solanum tuberosum]	EST492465 cSTS Solanum tuberosum cDNA clo
459	G2346	BG442540	1.00E-23	[Gossypium arboreum]	GA_Ea0017G06f Gossypium arboreum 7-10 d
459	G2346	AZ919034	1.90E-23	[Zea mays]	1006013G02.x3 1006 - RescueMu Grid G Zea mays geno
459	G2346	BE596165	2.70E-23	[Sorghum bicolor]	P11_50_D04.b1_A002 Pathogen induced 1 (P11)
459	G2346	AI443033	2.30E-22	[Glycine max]	sa31a08.y1 Gm-c1004 Glycine max cDNA clone GENO
459	G2346	BF482644	4.30E-22	[Triticum aestivum]	WHE2301-2304_A21_A212S Wheat pre-anthesis
459	G2346	gi5931643	6.20E-45	[Antirrhinum majus]	squamosa promoter binding protein-homol
459	G2346	gi5931786	4.20E-26	[Zea mays]	SBP-domain protein 5.
459	G2346	gi8468036	3.30E-14	[Oryza sativa]	Similar to Arabidopsis thaliana chromosome 2
459	G2346	gi9087308	8.30E-08	[Mitochondrion Beta vulgaris var. altissima]	orf102a.
285	G1354	BG128374	2.90E-58	[Lycopersicon esculentum]	EST474020 tomato shoot/meristem Lyc
285	G1354	BE202831	1.90E-56	[Medicago truncatula]	EST402853 KV1 Medicago truncatula cDNA
285	G1354	AI161918	6.60E-55	[Populus tremula x Populus tremuloides]	A009P50U Hybrid aspen
285	G1354	AB028186	1.20E-53	[Oryza sativa]	mRNA for OsNAC7 protein, complete cds.
285	G1354	BE060921	8.00E-50	[Hordeum vulgare]	HVSMEg0013N15f Hordeum vulgare pre-anthesis
285	G1354	AF402603	1.50E-42	[Phaseolus vulgaris]	NAC domain protein NAC2 mRNA, complete c
285	G1354	BE357920	1.60E-42	[Sorghum bicolor]	DG1_23_F03.b1_A002 Dark Grown 1 (DG1) Sorgh
285	G1354	PHRNANAM	3.60E-42	[Petunia x hybrida]	P.hybrida mRNA encoding NAM protein.
285	G1354	AW185617	5.30E-40	[Glycine max]	se80b05.y1 Gm-c1023 Glycine max cDNA clone GENO
285	G1354	gi6006373	4.50E-63	[Oryza sativa]	Similar to NAM like protein (AC005310).
285	G1354	gi15148914	2.30E-44	[Phaseolus vulgaris]	NAC domain protein NAC2.
285	G1354	gi14485513	3.50E-44	[Solanum tuberosum]	putative NAC domain protein.
285	G1354	gi1279640	5.90E-44	[Petunia x hybrida]	NAM.
285	G1354	gi6175246	5.20E-41	[Lycopersicon esculentum]	jasmonic acid 2.

Table 5

285	G1354	gi4218535	5.10E-39	[Triticum sp.]	GRAB1 protein.
285	G1354	gi6732158	5.10E-39	[Triticum monococcum]	unnamed protein product.
285	G1354	gi7716952	3.30E-35	[Medicago truncatula]	NAC1.
285	G1354	gi4996349	2.50E-26	[Nicotiana tabacum]	NAC-domain protein.
285	G1354	gi2982275	3.10E-14	[Picea mariana]	ATAF1-like protein.
119	G1063	BH700922	4.50E-90	[Brassica oleracea]	BOMMZ07TR BO 2_3_KB Brassica oleracea gen
119	G1063	BE451174	2.40E-41	[Lycopersicon esculentum]	EST402062 tomato root, plants pre-a
119	G1063	AW832545	2.00E-40	[Glycine max]	sm12e10.y1 Gm-c1027 Glycine max cDNA clone GENO
119	G1063	AP004693	5.90E-37	[Oryza sativa]	chromosome 8 clone P0461F06, *** SEQUENCING IN
119	G1063	AP004462	4.40E-32	[Oryza sativa (japonica cultivar-group)]	() chromosome 8 clo
119	G1063	AT002234	8.90E-32	[Brassica rapa subsp. pekinensis]	AT002234 Flower bud cDNA Br
119	G1063	BF263465	5.40E-25	[Hordeum vulgare]	HV_CEa0006N02f Hordeum vulgare seedling gre
119	G1063	BG557011	4.20E-22	[Sorghum bicolor]	EM1_41_E02.g1_A002 Embryo 1 (EM1) Sorghum b
119	G1063	BG842856	3.10E-21	[Zea mays]	MEST40-H05.T3 ISUM4-TN Zea mays cDNA clone MEST40-
119	G1063	BG559930	1.40E-18	[Sorghum propinquum]	RHIZ2_75_D09.g1_A003 Rhizome2 (RHIZ2) So
119	G1063	gi15528743	4.20E-26	[Oryza sativa]	contains EST C74560(E31855)-unknown protein.
119	G1063	gi6166283	8.10E-10	[Pinus taeda]	helix-loop-helix protein 1A.
119	G1063	gi11045087	8.80E-09	[Brassica napus]	putative protein.
119	G1063	gi10998404	7.10E-08	[Petunia x hybrida]	anthocyanin 1.
119	G1063	gi99441	2.60E-07	[Volvox carter]	sulfated surface glycoprotein 185 - Volvox
119	G1063	gi1142621	5.00E-07	[Phaseolus vulgaris]	phaseolin G-box binding protein PG2.
119	G1063	gi166428	8.10E-07	[Antirrhinum majus]	DEL.
119	G1063	gi1247386	9.50E-07	[Nicotiana glauca]	PRP2.
119	G1063	gi82091	1.00E-06	[Lycopersicon esculentum]	hydroxyproline-rich glycoprotein
119	G1063	gi1486263	1.40E-06	[Catharanthus roseus]	extensin.
129	G2143	BH650724	3.00E-88	[Brassica oleracea]	BOMIW43TR BO 2_3_KB Brassica oleracea gen
129	G2143	AW832545	1.50E-40	[Glycine max]	sm12e10.y1 Gm-c1027 Glycine max cDNA clone GENO
129	G2143	BE451174	3.50E-40	[Lycopersicon esculentum]	EST402062 tomato root, plants pre-a
129	G2143	AP004693	4.00E-38	[Oryza sativa]	chromosome 8 clone P0461F06, *** SEQUENCING IN
129	G2143	AP004584	6.30E-33	[Oryza sativa (japonica cultivar-group)]	() chromosome 8 clo
129	G2143	AT002234	3.00E-31	[Brassica rapa subsp. pekinensis]	AT002234 Flower bud cDNA Br
129	G2143	BF263465	2.90E-26	[Hordeum vulgare]	HV_CEa0006N02f Hordeum vulgare seedling gre
129	G2143	BG557011	2.60E-22	[Sorghum bicolor]	EM1_41_E02.g1_A002 Embryo 1 (EM1) Sorghum b
129	G2143	BG842856	3.50E-20	[Zea mays]	MEST40-H05.T3 ISUM4-TN Zea mays cDNA clone MEST40-
129	G2143	BG559930	6.10E-18	[Sorghum propinquum]	RHIZ2_75_D09.g1_A003 Rhizome2 (RHIZ2) So
129	G2143	gi15528743	5.50E-26	[Oryza sativa]	contains EST C74560(E31855)-unknown protein.

Table 5

129	G2143	gi1086538	7.60E-09	[Oryza rufipogon]	transcriptional activator Rb homolog.
129	G2143	gi6166283	1.10E-08	[Pinus taeda]	helix-loop-helix protein 1A.
129	G2143	gi1142621	4.60E-07	[Phaseolus vulgaris]	phaseolin G-box binding protein PG2.
129	G2143	gi3399777	5.20E-07	[Glycine max]	symbiotic ammonium transporter; nodulin.
129	G2143	gi5923912	6.10E-07	[Tulipa gesneriana]	bHLH transcription factor GBOF-1.
129	G2143	gi10998404	9.20E-07	[Petunia x hybrida]	anthocyanin 1.
129	G2143	gi4321762	5.20E-06	[Zea mays]	transcription factor MYC7E.
129	G2143	gi166428	6.00E-06	[Antirrhinum majus]	DEL.
129	G2143	gi527665	7.40E-06	[Sorghum bicolor]	myc-like regulatory R gene product.
133	G2557	BH511840	6.70E-62	[Brassica oleracea]	BOGRJ19TR BOGR Brassica oleracea genomic
133	G2557	BE347811	3.70E-46	[Glycine max]	sp05h10.y1 Gm-c1041 Glycine max cDNA clone GENO
133	G2557	AP003141	2.40E-33	[Oryza sativa]	genomic DNA, chromosome 1, PAC clone:P0002B05,
133	G2557	BF263465	3.00E-31	[Hordeum vulgare]	HV_CEa0006N02f Hordeum vulgare seedling gre
133	G2557	AT002234	6.60E-27	[Brassica rapa subsp. pekinensis]	AT002234 Flower bud cDNA Br
133	G2557	BG557011	6.40E-26	[Sorghum bicolor]	EM1_41_E02.g1_A002 Embryo 1 (EM1) Sorghum b
133	G2557	AP004462	7.90E-26	[Oryza sativa (japonica cultivar-group)]	() chromosome 8 clo
133	G2557	BE451174	3.90E-25	[Lycopersicon esculentum]	EST402062 tomato root, plants pre-a
133	G2557	BG842856	5.60E-22	[Zea mays]	MEST40-H05.T3 ISUM4-TN Zea mays cDNA clone MEST40-
133	G2557	BG559930	7.00E-14	[Sorghum propinquum]	RHIZ2_75_D09.g1_A003 Rhizome2 (RHIZ2) So
133	G2557	gi15289790	2.40E-36	[Oryza sativa]	contains EST C74560(E31855)-unknown protein.
133	G2557	gi3399777	2.60E-06	[Glycine max]	symbiotic ammonium transporter; nodulin.
133	G2557	gi4206118	1.10E-05	[Mesembryanthemum crystallinum]	transporter homolog.
133	G2557	gi6166283	1.30E-05	[Pinus taeda]	helix-loop-helix protein 1A.
133	G2557	gi527655	3.70E-05	[Pennisetum glaucum]	myc-like regulatory R gene product.
133	G2557	gi5923912	3.70E-05	[Tulipa gesneriana]	bHLH transcription factor GBOF-1.
133	G2557	gi527661	7.80E-05	[Phyllostachys acuta]	myc-like regulatory R gene product.
133	G2557	gi527665	9.50E-05	[Sorghum bicolor]	myc-like regulatory R gene product.
133	G2557	gi1086538	0.0001	[Oryza rufipogon]	transcriptional activator Rb homolog.
133	G2557	gi5669656	0.00013	[Lycopersicon esculentum]	ER33 protein.
697	G2430	BF632520	1.90E-14	[Medicago truncatula]	NF039A08DT1F1054 Drought Medicago trunc
697	G2430	AW396912	1.20E-13	[Glycine max]	sg64g09.y1 Gm-c1007 Glycine max cDNA clone GENO
697	G2430	D41804	4.50E-13	[Oryza sativa]	RICS4626A Rice shoot Oryza sativa cDNA, mRNA s
697	G2430	BE214029	2.60E-10	[Hordeum vulgare]	HV_CE00001P06f Hordeum vulgare seedling gre
697	G2430	AW564570	2.70E-10	[Sorghum bicolor]	LG1_296_E01.b1_A002 Light Grown 1 (LG1) Sor
697	G2430	BG129795	5.40E-10	[Lycopersicon esculentum]	EST475441 tomato shoot/meristem Lyc
697	G2430	AB060130	5.40E-09	[Zea mays]	ZmRR8 mRNA for response regulator 8, complete cds.

Table 5

697	G2430	BF587105	2.50E-05	[Sorghum propinquum]	FM1_32_C05.b1_A003 Floral-Induced Merist
697	G2430	AI163121	0.3	[Populus tremula x Populus tremuloides]	A033P70U Hybrid aspen
697	G2430	BG595628	0.46	[Solanum tuberosum]	EST494306 cSTS Solanum tuberosum cDNA clo
697	G2430	gi13661174	5.40E-18	[Zea mays]	response regulator 8.
697	G2430	gi15289981	0.028	[Oryza sativa]	hypothetical protein.
697	G2430	gi6942190	0.12	[Mesembryanthemum crystallinum]	CDPK substrate protein 1; C
697	G2430	gi4519671	0.2	[Nicotiana tabacum]	transfactor.
831	G1478	BF275913	1.50E-20	[Gossypium arboreum]	GA_EB0025C07f Gossypium arboreum 7-10 d
831	G1478	BG157399	6.50E-19	[Glycine max]	sab36g12.y1 Gm-c1026 Glycine max cDNA clone GEN
831	G1478	C95300	2.20E-10	[Citrus unshiu]	C95300 Citrus unshiu Miyagawa-wase maturation
831	G1478	AW034552	2.70E-10	[Lycopersicon esculentum]	EST278168 tomato callus, TAMU Lycop
831	G1478	BI070429	3.40E-10	[Populus tremula x Populus tremuloides]	C037P68U Populus sitra
831	G1478	AF016011	5.10E-09	[Brassica napus]	CONSTANS homolog (Bn9CON10) gene, complete c
831	G1478	BE598912	6.20E-09	[Sorghum bicolor]	PI1_84_H11.b1_A002 Pathogen induced 1 (PI1)
831	G1478	BG605313	6.80E-09	[Triticum aestivum]	WHE2331_C04_F07ZS Wheat pre-anthesis spik
831	G1478	BE558327	8.90E-09	[Hordeum vulgare]	HV_CEB0017D19f Hordeum vulgare seedling gre
831	G1478	BG647091	1.20E-08	[Medicago truncatula]	EST508710 HOGA Medicago truncatula cDNA
831	G1478	gi2895188	4.70E-11	[Brassica napus]	CONSTANS homolog.
831	G1478	gi3618308	1.50E-09	[Oryza sativa]	zinc finger protein.
831	G1478	gi11037308	4.70E-09	[Brassica nigra]	constans-like protein.
831	G1478	gi3341723	1.30E-08	[Raphanus sativus]	CONSTANS-like 1 protein.
831	G1478	gi4091806	1.50E-07	[Malus x domestica]	CONSTANS-like protein 2.
831	G1478	gi10946337	3.10E-07	[Ipomoea nil]	CONSTANS-like protein.
831	G1478	gi4557093	1.40E-05	[Pinus radiata]	zinc finger protein.
831	G1478	gi619312	0.9	[Capparis masaiikai]	mabinlin III B-chain=sweet protein mabi
831	G1478	gi4732091	1	[Zea mays]	bundle sheath defective protein 2.
831	G1478	gi4699629	1	[Nicotiana glauca]	Chain A, Putative Ancestral Protein Encod
579	G681	BG128147	6.80E-41	[Lycopersicon esculentum]	EST473793 tomato shoot/meristem Lyc
579	G681	BF054497	1.50E-39	[Solanum tuberosum]	EST439727 potato leaves and petioles Sola
579	G681	BE054276	8.40E-39	[Gossypium arboreum]	GA_Ea0002018f Gossypium arboreum 7-10 d
579	G681	BG269414	4.00E-38	[Mesembryanthemum crystallinum]	L0-3478T3 Ice plant Lambda Un
579	G681	BF620286	7.40E-38	[Hordeum vulgare]	HVSMEc0019F08f Hordeum vulgare seedling sho
579	G681	BE490032	1.00E-37	[Triticum aestivum]	WHE0364_C04_E08ZS Wheat cold-stressed see
579	G681	BI542536	1.40E-36	[Zea mays]	949021A03.y1 949 - Juvenile leaf and shoot cDNA fr
579	G681	BF425254	7.20E-36	[Glycine max]	su42c10.y1 Gm-c1068 Glycine max cDNA clone GENO
579	G681	AW672062	3.20E-34	[Sorghum bicolor]	LG1_354_G05.b1_A002 Light Grown 1 (LG1) Sor

Table 5

579	G681	BG448527	1.00E-33	[Medicago truncatula]	NF036F04RT1F1032 Developing root Medica
579	G681	gi13346188	9.10E-37	[Gossypium hirsutum]	GHMYB25.
579	G681	gi20563	6.30E-36	[Petunia x hybrida]	protein 1.
579	G681	gi485867	1.20E-34	[Antirrhinum majus]	mixta.
579	G681	gi2605617	1.70E-32	[Oryza sativa]	OSMYB1.
579	G681	gi1430846	2.00E-31	[Lycopersicon esculentum]	myb-related transcription factor.
579	G681	gi6651292	2.20E-30	[Pimpinella brachycarpa]	myb-related transcription factor.
579	G681	gi15042116	4.90E-30	[Zea mays subsp. parviglumis]	Cl protein.
579	G681	gi82730	6.10E-30	[Zea mays]	transforming protein (myb) homolog (clone Zm38)
579	G681	gi5139806	8.30E-30	[Glycine max]	GmMYB29A2.
579	G681	gi19055	1.10E-29	[Hordeum vulgare]	MybHv5.
611	G878	AF096299	6.20E-90	[Nicotiana tabacum]	DNA-binding protein 2 (WRKY2) mRNA, compl
611	G878	CUSLDB	1.80E-83	[Cucumis sativus]	SPF1-like DNA-binding protein mRNA, complet
611	G878	AF193802	3.50E-63	[Oryza sativa]	zinc finger transcription factor WRKY1 mRNA, c
611	G878	AX192162	2.20E-62	[Glycine max]	Sequence 9 from Patent WO0149840.
611	G878	IPBSPF1P	3.80E-58	[Ipomoea batatas]	Sweet potato mRNA for SPF1 protein, complet
611	G878	AFABF1	2.00E-56	[Avena fatua]	A. fatua mRNA for DNA-binding protein (clone ABF
611	G878	LES303343	7.20E-55	[Lycopersicon esculentum]	mRNA for hypothetical protein (ORF
611	G878	AX192164	4.00E-54	[Triticum aestivum]	Sequence 11 from Patent WO0149840.
611	G878	AF080595	2.10E-53	[Pimpinella brachycarpa]	zinc finger protein (ZFP1) mRNA, com
611	G878	PCU48831	2.30E-53	[Petroselinum crispum]	DNA-binding protein WRKY1 mRNA, comple
611	G878	gi4322940	3.30E-128	[Nicotiana tabacum]	DNA-binding protein 2.
611	G878	gi927025	1.10E-109	[Cucumis sativus]	SPF1-like DNA-binding protein.
611	G878	gi6689916	1.50E-74	[Oryza sativa]	zinc finger transcription factor WRKY1.
611	G878	gi484261	1.10E-66	[Ipomoea batatas]	SPF1 protein.
611	G878	gi1159877	2.30E-63	[Avena fatua]	DNA-binding protein.
611	G878	gi13620227	4.60E-63	[Lycopersicon esculentum]	hypothetical protein.
611	G878	gi5917653	1.70E-56	[Petroselinum crispum]	zinc-finger type transcription facto
611	G878	gi4894965	5.00E-56	[Avena sativa]	DNA-binding protein WRKY1.
611	G878	gi3420906	8.70E-56	[Pimpinella brachycarpa]	zinc finger protein; WRKY1.
611	G878	gi13620168	4.20E-22	[Capsella rubella]	hypothetical protein.
47	G374	AP004457	1.20E-73	[Oryza sativa (japonica cultivar-group)]	() chromosome 8 clo
47	G374	AP004693	1.90E-73	[Oryza sativa]	chromosome 8 clone P0461F06, *** SEQUENCING IN
47	G374	BH552835	1.30E-62	[Brassica oleracea]	BOHGT56TR BOHG Brassica oleracea genomic
47	G374	BG128229	6.50E-55	[Lycopersicon esculentum]	EST473875 tomato shoot/meristem Lyc
47	G374	BG646959	3.20E-46	[Medicago truncatula]	EST508578 HOGA Medicago truncatula cDNA

Table 5

47	G374	BG890162	8.70E-41	[Solanum tuberosum]	EST516013 cSTD Solanum tuberosum cDNA clo
47	G374	AW179366	6.00E-38	[Zea mays]	618046G06.y1 618 - Inbred Tassel cDNA Library Zea
47	G374	BF473206	1.50E-32	[Triticum aestivum]	WHE0922 G12 M24ZS Wheat 5-15 DAP spike cD
47	G374	AW761011	2.90E-29	[Glycine max]	sl61g11.y1 Gm-c1027 Glycine max cDNA clone GENO
47	G374	AJ436050	1.50E-27	[Hordeum vulgare]	AJ436050 S00007 Hordeum vulgare cDNA clone
47	G374	gi422012	0.8	[Sorghum bicolor]	lipid transfer protein - sorghum (fragmen
47	G374	gi1827893	1	[Zea mays]	Maize Nonspecific Lipid Transfer Protein Complex

Traits of interest

Examples of some of the traits that may be desirable in plants, and that may be provided by transforming the plants with the presently disclosed sequences, are listed in Table 6.

Table 6. Genes, traits and utilities that affect plant characteristics

<u>Trait Category</u>	<u>Traits</u>	<u>Transcription factor genes that impact traits</u>	<u>Utility</u> <u>Gene effect on:</u>
Resistance and tolerance	Salt stress resistance	G22; G196; G226; G303; G312; G325; G353; G482; G545; G801; G867; G884; G922; G926; G1452; G1794; G1820; G1836; G1843; G1863; G2053; G2110; G2140; G2153; G2379; G2701; G2713; G2719; G2789	Germination rate, survivability, yield; extended growth range
	Osmotic stress resistance	G47; G175; G188; G303; G325; G353; G489; G502; G526; G921; G922; G926; G1069; G1089; G1452; G1794; G1930; G2140; G2153; G2379; G2701; G2719; G2789;	Germination rate, survivability, yield
	Cold stress resistance; cold germination	G256; G394; G664; G864; G1322; G2130	Germination, growth, earlier planting
	Tolerance to freezing	G303; G325; G353; G720; G912; G913; G1794; G2053; G2140; G2153; G2379; G2701; G2719; G2789	Survivability, yield, appearance, extended range
	Heat stress resistance	G3; G464; G682; G864; G964;	Germination,

		G1305; G1645; G2130 G2430	growth, later planting
	Drought, low humidity resistance	G303; G325; G353; G720; G912; G926; G1452; G1794; G1820; G1843; G2053; G2140; G2153; G2379; G2583; G2701; G2719; G2789	Survivability, yield, extended range
	Radiation resistance	G1052	Survivability, vigor, appearance
	Decreased herbicide sensitivity	G343; G2133; G2517	Resistant to increased herbicide use
	Increased herbicide sensitivity	G374; G877; G1519	Use as a herbicide target
	Oxidative stress	G477; G789; G1807; G2133; G2517	Improved yield, appearance, reduced senescence
	Light response	G183; G354; G375; G1062; G1322; G1331; G1488; G1494; G1521; G1786; G1794; G2144; G2555;	Germination, growth, development, flowering time
Development, morphology	Overall plant architecture	G24; G27; G31; G33; G47; G147; G156; G160; G182; G187; G195; G196; G211; G221; G237; G280; G342; G352; G357; G358; G360; G362; G364; G365; G367; G373; G377; G396; G431; G447; G479; G546; G546; G551; G578; G580; G596; G615; G617; G620; G625;	Vascular tissues, lignin content; cell wall content; appearance

		G638; G658; G716; G725; G727; G730; G740; G770; G858; G865; G869; G872; G904; G910; G912; G920; G939; G963; G977; G979; G987; G988; G993; G1007; G1010; G1014; G1035; G1046; G1049; G1062; G1069; G1070; G1076; G1089; G1093; G1127; G1131; G1145; G1229; G1246; G1304; G1318; G1320; G1330; G1331; G1352; G1354; G1360; G1364; G1379; G1384; G1399; G1415; G1417; G1442; G1453; G1454; G1459; G1460; G1471; G1475; G1477; G1487; G1487; G1492; G1499; G1499; G1531; G1540; G1543; G1543; G1544; G1548; G1584; G1587; G1588; G1589; G1636; G1642; G1747; G1749; G1749; G1751; G1752; G1763; G1766; G1767; G1778; G1789; G1790; G1791; G1793; G1794; G1795; G1800; G1806; G1811; G1835; G1836; G1838; G1839; G1843; G1853; G1855; G1865; G1881; G1882; G1883; G1884; G1891; G1896; G1898; G1902; G1904; G1906; G1913; G1914; G1925; G1929; G1930; G1954; G1958; G1965; G1976; G2057; G2107; G2133; G2134; G2151; G2154; G2157; G2181;	
--	--	--	--

		G2290; G2299; G2340; G2340; G2346; G2373; G2376; G2424; G2465; G2505; G2509; G2512; G2513; G2519; G2520; G2533; G2534; G2573; G2589; G2687; G2720; G2787; G2789; G2893	
	Size: increased stature	G189; G1073; G1435; G2430	
	Size: reduced stature or dwarfism	G3; G5; G21; G23; G39; G165; G184; G194; G258; G280; G340; G343; G353; G354; G362; G363; G370; G385; G396; G439; G440; G447; G450; G550; G557; G599; G636; G652; G670; G671; G674; G729; G760; G804; G831; G864; G884; G898; G900; G912; G913; G922; G932; G937; G939; G960; G962; G977; G991; G1000; G1008; G1020; G1023; G1053; G1067; G1075; G1137; G1181; G1198; G1228; G1266; G1267; G1275; G1277; G1309; G1311; G1314; G1317; G1322; G1323; G1326; G1332; G1334; G1367; G1381; G1382; G1386; G1421; G1488; G1494; G1537; G1545; G1560; G1586; G1641; G1652; G1655; G1671; G1750; G1756; G1757; G1782; G1786; G1794; G1839; G1845; G1879; G1886; G1888; G1933; G1939; G1943; G1944; G2011; G2094; G2115;	Ornamental; small stature provides wind resistance; creation of dwarf varieties

		G2130; G2132; G2144; G2145; G2147; G2156; G2294; G2313; G2344; G2431; G2510; G2517; G2521; G2893; G2893	
	Fruit size and number	G362	Biomass, yield, cotton boll fiber density
	Flower structure, inflorescence	G47; G259; G353; G354; G671; G732; G988; G1000; G1063; G1140; G1326; G1449; G1543; G1560; G1587; G1645; G1947; G2108; G2143; G2893	Ornamental horticulture; production of saffron or other edible flowers
	Number and development of trichomes	G225; G226; G247; G362; G585; G634; G676; G682; G1014; G1332; G1452; G1795; G2105	Resistance to pests and desiccation; essential oil production
	Seed size, color, and number	G156; G450; G584; G652; G668; G858; G979; G1040; G1062; G1145; G1255; G1494; G1531; G1534; G1594; G2105; G2114;	Yield
	Root development, modifications	G9; G1482; G1534; G1794; G1852; G2053; G2136; G2140	
	Modifications to root hairs	G225; G226	Nutrient, water uptake, pathogen resistance
	Apical dominance	G559; G732; G1255; G1275; G1411; G1488; G1635; G2452; G2509	Ornamental horticulture
	Branching patterns	G568; G988; G1548	Ornamental horticulture, knot reduction, improved

			windscreen
	Leaf shape, color, modifications	G375; G377; G428; G438; G447; G464; G557; G577; G599; G635; G671; G674; G736; G804; G903; G977; G921; G922; G1038; G1063; G1067; G1073; G1075; G1146; G1152; G1198; G1267; G1269; G1452; G1484; G1586; G1594; G1767; G1786; G1792; G1886; G2059; G2094; G2105; G2113; G2117; G2143; G2144; G2431; G2452; G2465; G2587; G2583; G2724;	Appealing shape or shiny leaves for ornamental agriculture, increased biomass or photosynthesis
	Silique	G1134	Ornamental
	Stem morphology	G47; G438; G671; G748; G988; G1000	Ornamental; digestibility
	Shoot modifications	G390; G391	Ornamental stem bifurcations
Disease, Pathogen Resistance	Bacterial	G211; G347; G367; G418; G525; G545; G578; G1049	Yield, appearance, survivability, extended range
	Fungal	G19; G28; G28; G28; G147; G188; G207; G211; G237; G248; G278; G347; G367; G371; G378; G409; G477; G545; G545; G558; G569; G578; G591; G594; G616; G789; G805; G812; G865; G869; G872; G881; G896; G940; G1047; G1049; G1064; G1084; G1196; G1255; G1266;	Yield, appearance, survivability, extended range

		G1363; G1514; G1756; G1792; G1792; G1792; G1792; G1880; G1919; G1919; G1927; G1927; G1936; G1936; G1950; G2069; G2130; G2380; G2380; G2555	
Nutrients	Increased tolerance to nitrogen-limited soils	G225; G226; G1792	
	Increased tolerance to phosphate-limited soils	G419; G545; G561; G1946	
	Increased tolerance to potassium-limited soils	G561; G911	
Hormonal	Hormone sensitivity	G12; G546; G926; G760; G913; G926; G1062; G1069; G1095; G1134; G1330; G1452; G1666; G1820; G2140; G2789	Seed dormancy, drought tolerance; plant form, fruit ripening
Seed biochemistry	Production of seed prenyl lipids, including tocopherol	G214; G259; G490; G652; G748; G883; G1052; G1328; G1930; G2509; G2520	Antioxidant activity, vitamin E
	Production of seed sterols	G20	Precursors for human steroid hormones; cholesterol modulators
	Production of seed glucosinolates	G353; G484; G674; G1272; G1506; G1897; G1946; G2113; G2117; G2155; G2290; G2340	Defense against insects; putative anticancer activity; undesirable in

			animal feeds
	Modified seed oil content	G162; G162; G180; G192; G241; G265; G286; G291; G427; G509; G519; G561; G567; G590; G818; G849; G892; G961; G974; G1063; G1143; G1190; G1198; G1226; G1229; G1323; G1451; G1471; G1478; G1496; G1526; G1543; G1640; G1644; G1646; G1672; G1677; G1750; G1765; G1777; G1793; G1838; G1902; G1946; G1948; G2059; G2123; G2138; G2139; G2343; G2792; G2830	Vegetable oil production; increased caloric value for animal feeds; lutein content
	Modified seed oil composition	G217; G504; G622; G778; G791; G861; G869; G938; G965; G1417; G2192	Heat stability, digestibility of seed oils
	Modified seed protein content	G162; G226; G241; G371; G427; G509; G567; G597; G732; G849; G865; G892; G963; G988; G1323; G1323; G1419; G1478; G1488; G1634; G1637; G1641; G1644; G1652; G1677; G1777; G1777; G1818; G1820; G1903; G1909; G1946; G1946; G1958; G2059; G2117; G2417; G2509	Reduced caloric value for humans
Leaf biochemistry	Production of flavonoids	G1666*	Ornamental pigment production; pathogen resistance; health

			benefits
	Production of leaf glucosinolates	G264; G353; G484; G652; G674; G681; G1069; G1198; G1322; G1421; G1657; G1794; G1897; G1946; G2115; G2117; G2144; G2155; G2155; G2340; G2512; G2520; G2552	Defense against insects; putative anticancer activity; undesirable in animal feeds
	Production of diterpenes	G229	Induction of enzymes involved in alkaloid biosynthesis
	Production of anthocyanin	G546	Ornamental pigment
	Production of leaf phytosterols, inc. stigmastanol, campesterol	G561; G2131; G2424	Precursors for human steroid hormones; cholesterol modulators
	Leaf fatty acid composition	G214; G377; G861; G962; G975; G987; G1266; G1337; G1399; G1465; G1512; G2136; G2147; G2192	Nutritional value; increase in waxes for disease resistance
	Production of leaf prenyl lipids, including tocopherol	G214; G259; G280; G652; G987; G1543; G2509; G2520	Antioxidant activity, vitamin E
Biochemistry, general	Production of miscellaneous secondary metabolites	G229; G663	
	Sugar, starch, hemicellulose composition,	G158; G211; G211; G237; G242; G274; G598; G1012; G1266; G1309; G1309; G1641; G1765; G1865; G2094; G2094;	Food digestibility, hemicellulose & pectin content; fiber content; plant

		G2589; G2589	tensile strength, wood quality, pathogen resistance, pulp production; tuber starch content
Sugar sensing	Plant response to sugars	G26; G38; G43; G207; G218; G241; G254; G263; G308; G536; G567; G567; G680; G867; G912; G956; G996; G1068; G1225; G1314; G1314; G1337; G1759; G1804; G2153; G2379	Photosynthetic rate, carbohydrate accumulation, biomass production, source-sink relationships, senescence
Growth, Reproduction	Plant growth rate and development	G447; G617; G674; G730; G917; G937; G1035; G1046; G1131; G1425; G1452; G1459; G1492; G1589; G1652; G1879; G1943; G2430; G2431; G2465; G2521	Faster growth, increased biomass or yield, improved appearance; delay in bolting
	Embryo development	G167	
	Seed germination rate	G979; G1792; G2130	Yield
	Plant, seedling vigor	G561; G2346	Survivability, yield
	Senescence; cell death	G571; G636; G878; G1050; G1463; G1749; G1944; G2130; G2155; G2340; G2383	Yield, appearance; response to pathogens;
	Modified fertility	G39; G340; G439; G470; G559; G615; G652; G671; G779; G962; G977; G988; G1000; G1063; G1067; G1075;	Prevents or minimizes escape of the pollen of GMOs

		G1266; G1311; G1321; G1326; G1367; G1386; G1421; G1453; G1471; G1453; G1560; G1594; G1635; G1750; G1947; G2011; G2094; G2113; G2115; G2130; G2143; G2147; G2294; G2510; G2893	
	Early flowering	G147; G157; G180; G183; G183; G184; G185; G208; G227; G294; G390; G390; G390; G391; G391; G427; G427; G490; G565; G590; G592; G720; G789; G865; G898; G898; G989; G989; G1037; G1037; G1142; G1225; G1225; G1226; G1242; G1305; G1305; G1380; G1380; G1480; G1480; G1488; G1494; G1545; G1545; G1649; G1706; G1760; G1767; G1767; G1820; G1841; G1841; G1842; G1843; G1843; G1946; G1946; G2010; G2030; G2030; G2144; G2144; G2295; G2295; G2347; G2348; G2348; G2373; G2373; G2509; G2509; G2555; G2555	Faster generation time; synchrony of flowering; potential for introducing new traits to single variety
	Delayed flowering	G8; G47; G192; G214; G234; G361; G362; G562; G568; G571; G591; G680; G736; G748; G859; G878; G910; G912; G913; G971; G994; G1051; G1052; G1073; G1079; G1335; G1435; G1452; G1478;	Delayed time to pollen production of GMO plants; synchrony of flowering; increased yield

		G1789; G1804; G1865; G1865; G1895; G1900; G2007; G2133; G2155; G2291; G2465	
	Extended flowering phase	G1947	
	Flower and leaf development	G259; G353; G377; G580; G638 G652; G858; G869; G917; G922; G932; G1063; G1075; G1140; G1425; G1452; G1499; G1548; G1645; G1865; G1897; G1933; G2094; G2124; G2140; G2143; G2535; G2557	Ornamental applications; decreased fertility
	Flower abscission	G1897	Ornamental: longer retention of flowers

* When co-expressed with G669 and G663

Significance of modified plant traits

Currently, the existence of a series of maturity groups for different latitudes represents a major barrier to the introduction of new valuable traits. Any trait (e.g. disease resistance) has to be bred into each of the different maturity groups separately, a laborious and costly exercise. The availability of single strain, which could be grown at any latitude, would therefore greatly increase the potential for introducing new traits to crop species such as soybean and cotton.

For many of the traits, listed in Table 6 and below, that may be conferred to plants, a single transcription factor gene may be used to increase or decrease, advance or delay, or improve or prove deleterious to a given trait. For example, overexpression of a transcription factor gene that naturally occurs in a plant may cause early flowering relative to non-transformed or wild-type plants. By knocking out the gene, or suppressing the gene (with, for example, antisense suppression) the plant may experience delayed flowering. Similarly, overexpressing or suppressing one or more genes can impart significant differences in production of plant products,

such as different fatty acid ratios. Thus, suppressing a gene that causes a plant to be more sensitive to cold may improve a plant's tolerance of cold.

Salt stress resistance. Soil salinity is one of the more important variables that determines where a plant may thrive. Salinity is especially important for the successful cultivation of crop plants, particular in many parts of the world that have naturally high soil salt concentrations, or where the soil has been over-utilized. Thus, presently disclosed transcription factor genes that provide increased salt tolerance during germination, the seedling stage, and throughout a plant's life cycle would find particular value for imparting survivability and yield in areas where a particular crop would not normally prosper.

Osmotic stress resistance. Presently disclosed transcription factor genes that confer resistance to osmotic stress may increase germination rate under adverse conditions, which could impact survivability and yield of seeds and plants.

Cold stress resistance. The potential utility of presently disclosed transcription factor genes that increase tolerance to cold is to confer better germination and growth in cold conditions. The germination of many crops is very sensitive to cold temperatures. Genes that would allow germination and seedling vigor in the cold would have highly significant utility in allowing seeds to be planted earlier in the season with a high rate of survivability. Transcription factor genes that confer better survivability in cooler climates allow a grower to move up planting time in the spring and extend the growing season further into autumn for higher crop yields.

Tolerance to freezing. The presently disclosed transcription factor genes that impart tolerance to freezing conditions are useful for enhancing the survivability and appearance of plants conditions or conditions that would otherwise cause extensive cellular damage. Thus, germination of seeds and survival may take place at temperatures significantly below that of the mean temperature required for germination of seeds and survival of non-transformed plants. As with salt tolerance, this has the added benefit of increasing the potential range of a crop plant into regions in which it would otherwise succumb. Cold tolerant transformed plants may also be

planted earlier in the spring or later in autumn, with greater success than with non-transformed plants.

Heat stress tolerance. The germination of many crops is also sensitive to high temperatures. Presently disclosed transcription factor genes that provide increased heat tolerance are generally useful in producing plants that germinate and grow in hot conditions, may find particular use for crops that are planted late in the season, or extend the range of a plant by allowing growth in relatively hot climates.

Drought, low humidity tolerance. Strategies that allow plants to survive in low water conditions may include, for example, reduced surface area or surface oil or wax production. A number of presently disclosed transcription factor genes increase a plant's tolerance to low water conditions and provide the benefits of improved survivability, increased yield and an extended geographic and temporal planting range.

Radiation resistance. Presently disclosed transcription factor genes have been shown to increase lutein production. Lutein, like other xanthophylls such as zeaxanthin and violaxanthin, are important in the protection of plants against the damaging effects of excessive light. Lutein contributes, directly or indirectly, to the rapid rise of non-photochemical quenching in plants exposed to high light. Increased tolerance of field plants to visible and ultraviolet light impacts survivability and vigor, particularly for recent transplants. Also affected are the yield and appearance of harvested plants or plant parts. Crop plants engineered with presently disclosed transcription factor genes that cause the plant to produce higher levels of lutein therefore would have improved photoprotection, leading to less oxidative damage and increase vigor, survivability and higher yields under high light and ultraviolet light conditions.

Decreased herbicide sensitivity. Presently disclosed transcription factor genes that confer resistance or tolerance to herbicides (e.g., glyphosate) may find use in providing means to increase herbicide applications without detriment to desirable plants. This would allow for the increased use of a particular herbicide in a local

environment, with the effect of increased detriment to undesirable species and less harm to transgenic, desirable cultivars.

Increased herbicide sensitivity. Knockouts of a number of the presently disclosed transcription factor genes have been shown to be lethal to developing embryos. Thus, these genes are potentially useful as herbicide targets.

Oxidative stress. In plants, as in all living things, abiotic and biotic stresses induce the formation of oxygen radicals, including superoxide and peroxide radicals. This has the effect of accelerating senescence, particularly in leaves, with the resulting loss of yield and adverse effect on appearance. Generally, plants that have the highest level of defense mechanisms, such as, for example, polyunsaturated moieties of membrane lipids, are most likely to thrive under conditions that introduce oxidative stress (e.g., high light, ozone, water deficit, particularly in combination). Introduction of the presently disclosed transcription factor genes that increase the level of oxidative stress defense mechanisms would provide beneficial effects on the yield and appearance of plants. One specific oxidizing agent, ozone, has been shown to cause significant foliar injury, which impacts yield and appearance of crop and ornamental plants. In addition to reduced foliar injury that would be found in ozone resistant plant created by transforming plants with some of the presently disclosed transcription factor genes, the latter have also been shown to have increased chlorophyll fluorescence (Yu-Sen Chang et al. Bot. Bull. Acad. Sin. (2001) 42: 265-272).

Heavy metal tolerance. Heavy metals such as lead, mercury, arsenic, chromium and others may have a significant adverse impact on plant respiration. Plants that have been transformed with presently disclosed transcription factor genes that confer improved resistance to heavy metals, through, for example, sequestering or reduced uptake of the metals will show improved vigor and yield in soils with relatively high concentrations of these elements. Conversely, transgenic transcription factors may also be introduced into plants to confer an increase in heavy metal uptake, which may benefit efforts to clean up contaminated soils.

Light response. Presently disclosed transcription factor genes that modify a plant's response to light may be useful for modifying a plant's growth or

development, for example, photomorphogenesis in poor light, or accelerating flowering time in response to various light intensities, quality or duration to which a non-transformed plant would not similarly respond. Examples of such responses that have been demonstrated include leaf number and arrangement, and early flower bud appearances.

Overall plant architecture. Several presently disclosed transcription factor genes have been introduced into plants to alter numerous aspects of the plant's morphology. For example, it has been demonstrated that a number of transcription factors may be used to manipulate branching, such as the means to modify lateral branching, a possible application in the forestry industry. Transgenic plants have also been produced that have altered cell wall content, lignin production, flower organ number, or overall shape of the plants. Presently disclosed transcription factor genes transformed into plants may be used to affect plant morphology by increasing or decreasing internode distance, both of which may be advantageous under different circumstances. For example, for fast growth of woody plants to provide more biomass, or fewer knots, increased internode distances are generally desirable. For improved wind screening of shrubs or trees, or harvesting characteristics of, for example, members of the Gramineae family, decreased internode distance may be advantageous. These modifications would also prove useful in the ornamental horticulture industry for the creation of unique phenotypic characteristics of ornamental plants.

Increased stature. For some ornamental plants, the ability to provide larger varieties may be highly desirable. For many plants, including fruit-bearing trees or trees and shrubs that serve as view or wind screens, increased stature provides obvious benefits. Crop species may also produce higher yields on larger cultivars.

Reduced stature or dwarfism. Presently disclosed transcription factor genes that decrease plant stature can be used to produce plants that are more resistant to damage by wind and rain, or more resistant to heat or low humidity or water deficit. Dwarf plants are also of significant interest to the ornamental horticulture industry, and particularly for home garden applications for which space availability may be limited.

Fruit size and number. Introduction of presently disclosed transcription factor genes that affect fruit size will have desirable impacts on fruit size and number, which may comprise increases in yield for fruit crops, or reduced fruit yield, such as when vegetative growth is preferred (e.g., with bushy ornamentals, or where fruit is undesirable, as with ornamental olive trees).

Flower structure, inflorescence, and development. Presently disclosed transgenic transcription factors have been used to create plants with larger flowers or arrangements of flowers that are distinct from wild-type or non-transformed cultivars. This would likely have the most value for the ornamental horticulture industry, where larger flowers or interesting presentations generally are preferred and command the highest prices. Flower structure may have advantageous effects on fertility, and could be used, for example, to decrease fertility by the absence, reduction or screening of reproductive components. One interesting application for manipulation of flower structure, for example, by introduced transcription factors could be in the increased production of edible flowers or flower parts, including saffron, which is derived from the stigmas of *Crocus sativus*.

Number and development of trichomes. Several presently disclosed transcription factor genes have been used to modify trichome number and amount of trichome products in plants. Trichome glands on the surface of many higher plants produce and secrete exudates that give protection from the elements and pests such as insects, microbes and herbivores. These exudates may physically immobilize insects and spores, may be insecticidal or ant-microbial or they may act as allergens or irritants to protect against herbivores. Trichomes have also been suggested to decrease transpiration by decreasing leaf surface air flow, and by exuding chemicals that protect the leaf from the sun.

Seed size, color and number. The introduction of presently disclosed transcription factor genes into plants that alter the size or number of seeds may have a significant impact on yield, both when the product is the seed itself, or when biomass of the vegetative portion of the plant is increased by reducing seed production. In the case of fruit products, it is often advantageous to modify a plant to have reduced size

or number of seeds relative to non-transformed plants to provide seedless or varieties with reduced numbers or smaller seeds. Presently disclosed transcription factor genes have also been shown to affect seed size, including the development of larger seeds. Seed size, in addition to seed coat integrity, thickness and permeability, seed water content and by a number of other components including antioxidants and oligosaccharides, may affect seed longevity in storage. This would be an important utility when the seed of a plant is the harvested crops, as with, for example, peas, beans, nuts, etc. Presently disclosed transcription factor genes have also been used to modify seed color, which could provide added appeal to a seed product.

Root development, modifications. By modifying the structure or development of roots by transforming into a plant one or more of the presently disclosed transcription factor genes, plants may be produced that have the capacity to thrive in otherwise unproductive soils. For example, grape roots that extend further into rocky soils, or that remain viable in waterlogged soils, would increase the effective planting range of the crop. It may be advantageous to manipulate a plant to produce short roots, as when a soil in which the plant will be growing is occasionally flooded, or when pathogenic fungi or disease-causing nematodes are prevalent.

Modifications to root hairs. Presently disclosed transcription factor genes that increase root hair length or number potentially could be used to increase root growth or vigor, which might in turn allow better plant growth under adverse conditions such as limited nutrient or water availability.

Apical dominance. The modified expression of presently disclosed transcription factors that control apical dominance could be used in ornamental horticulture, for example, to modify plant architecture.

Branching patterns. Several presently disclosed transcription factor genes have been used to manipulate branching, which could provide benefits in the forestry industry. For example, reduction in the formation of lateral branches could reduce knot formation. Conversely, increasing the number of lateral branches could provide utility when a plant is used as a windscreen, or may also provide ornamental advantages.

Leaf shape, color and modifications. It has been demonstrated in laboratory experiments that overexpression of some of the presently disclosed transcription factors produced marked effects on leaf development. At early stages of growth, these transgenic seedlings developed narrow, upward pointing leaves with long petioles, possibly indicating a disruption in circadian-clock controlled processes or nyctinastic movements. Other transcription factor genes can be used to increase plant biomass; large size would be useful in crops where the vegetative portion of the plant is the marketable portion.

Siliques. Genes that later silique conformation in brassicates may be used to modify fruit ripening processes in brassicates and other plants, which may positively affect seed or fruit quality.

Stem morphology and shoot modifications. Laboratory studies have demonstrated that introducing several of the presently disclosed transcription factor genes into plants can cause stem bifurcations in shoots, in which the shoot meristems split to form two or three separate shoots. This unique appearance would be desirable in ornamental applications.

Diseases, pathogens and pests. A number of the presently disclosed transcription factor genes have been shown to or are likely to confer resistance to various plant diseases, pathogens and pests. The offending organisms include fungal pathogens *Fusarium oxysporum*, *Botrytis cinerea*, *Sclerotinia sclerotiorum*, and *Erysiphe orontii*. Bacterial pathogens to which resistance may be conferred include *Pseudomonas syringae*. Other problem organisms may potentially include nematodes, mollicutes, parasites, or herbivorous arthropods. In each case, one or more transformed transcription factor genes may provide some benefit to the plant to help prevent or overcome infestation. The mechanisms by which the transcription factors work could include increasing surface waxes or oils, surface thickness, local senescence, or the activation of signal transduction pathways that regulate plant defense in response to attacks by herbivorous pests (including, for example, protease inhibitors).

Increased tolerance of plants to nutrient-limited soils. Presently disclosed transcription factor genes introduced into plants may provide the means to improve uptake of essential nutrients, including nitrogenous compounds, phosphates, potassium, and trace minerals. The effect of these modifications is to increase the seedling germination and range of ornamental and crop plants. The utilities of presently disclosed transcription factor genes conferring tolerance to conditions of low nutrients also include cost savings to the grower by reducing the amounts of fertilizer needed, environmental benefits of reduced fertilizer runoff, and improved yield and stress tolerance. In addition, this gene could be used to alter seed protein amounts and/or composition that could impact yield as well as the nutritional value and production of various food products.

Hormone sensitivity. One or more of the presently disclosed transcription factor genes have been shown to affect plant abscisic acid (ABA) sensitivity. This plant hormone is likely the most important hormone in mediating the adaptation of a plant to stress. For example, ABA mediates conversion of apical meristems into dormant buds. In response to increasingly cold conditions, the newly developing leaves growing above the meristem become converted into stiff bud scales that closely wrap the meristem and protect it from mechanical damage during winter. ABA in the bud also enforces dormancy; during premature warm spells, the buds are inhibited from sprouting. Bud dormancy is eliminated after either a prolonged cold period of cold or a significant number of lengthening days. Thus, by affecting ABA sensitivity, introduced transcription factor genes may affect cold sensitivity and survivability. ABA is also important in protecting plants from drought tolerance.

Several other of the present transcription factor genes have been used to manipulate ethylene signal transduction and response pathways. These genes can thus be used to manipulate the processes influenced by ethylene, such as seed germination or fruit ripening, and to improve seed or fruit quality.

Production of seed and leaf prenyl lipids, including tocopherol. Prenyl lipids play a role in anchoring proteins in membranes or membranous organelles. Thus modifying the prenyl lipid content of seeds and leaves could affect membrane integrity and function. A number of presently disclosed transcription factor genes

have been shown to modify the tocopherol composition of plants. Tocopherols have both anti-oxidant and vitamin E activity.

Production of seed and leaf phytosterols: Presently disclosed transcription factor genes that modify levels of phytosterols in plants may have at least two utilities. First, phytosterols are an important source of precursors for the manufacture of human steroid hormones. Thus, regulation of transcription factor expression or activity could lead to elevated levels of important human steroid precursors for steroid semi-synthesis. For example, transcription factors that cause elevated levels of campesterol in leaves, or sitosterols and stigmasterols in seed crops, would be useful for this purpose. Phytosterols and their hydrogenated derivatives phytostanols also have proven cholesterol-lowering properties, and transcription factor genes that modify the expression of these compounds in plants would thus provide health benefits.

Production of seed and leaf glucosinolates. Some glucosinolates have anti-cancer activity; thus, increasing the levels or composition of these compounds by introducing several of the presently disclosed transcription factors might be of interest from a nutraceutical standpoint. (3) Glucosinolates form part of a plants natural defense against insects. Modification of glucosinolate composition or quantity could therefore afford increased protection from predators. Furthermore, in edible crops, tissue specific promoters might be used to ensure that these compounds accumulate specifically in tissues, such as the epidermis, which are not taken for consumption.

Modified seed oil content. The composition of seeds, particularly with respect to seed oil amounts and/or composition, is very important for the nutritional value and production of various food and feed products. Several of the presently disclosed transcription factor genes in seed lipid saturation that alter seed oil content could be used to improve the heat stability of oils or to improve the nutritional quality of seed oil, by, for example, reducing the number of calories in seed, increasing the number of calories in animal feeds, or altering the ratio of saturated to unsaturated lipids comprising the oils.

Seed and leaf fatty acid composition. A number of the presently disclosed transcription factor genes have been shown to alter the fatty acid composition in plants, and seeds in particular. This modification may find particular value for improving the nutritional value of, for example, seeds or whole plants. Dietary fatty acids ratios have been shown to have an effect on, for example, bone integrity and remodeling (see, for example, Weiler, H.A., *Pediatr Res* (2000) 47:5 692-697). The ratio of dietary fatty acids may alter the precursor pools of long-chain polyunsaturated fatty acids that serve as precursors for prostaglandin synthesis. In mammalian connective tissue, prostaglandins serve as important signals regulating the balance between resorption and formation in bone and cartilage. Thus dietary fatty acid ratios altered in seeds may affect the etiology and outcome of bone loss.

Modified seed protein content. As with seed oils, the composition of seeds, particularly with respect to protein amounts and/or composition, is very important for the nutritional value and production of various food and feed products. A number of the presently disclosed transcription factor genes modify the protein concentrations in seeds would provide nutritional benefits, and may be used to prolong storage, increase seed pest or disease resistance, or modify germination rates.

Production of flavonoids in leaves and other plant parts. Expression of presently disclosed transcription factor genes that increase flavonoid production in plants, including anthocyanins and condensed tannins, may be used to alter in pigment production for horticultural purposes, and possibly increasing stress resistance. Flavonoids have antimicrobial activity and could be used to engineer pathogen resistance. Several flavonoid compounds have health promoting effects such as the inhibition of tumor growth and cancer, prevention of bone loss and the prevention of the oxidation of lipids. Increasing levels of condensed tannins, whose biosynthetic pathway is shared with anthocyanin biosynthesis, in forage legumes is an important agronomic trait because they prevent pasture bloat by collapsing protein foams within the rumen. For a review on the utilities of flavonoids and their derivatives, refer to Dixon et al. (1999) *Trends Plant Sci.* 4:394-400.

Production of diterpenes in leaves and other plant parts. Depending on the plant species, varying amounts of diverse secondary biochemicals (often lipophilic

terpenes) are produced and exuded or volatilized by trichomes. These exotic secondary biochemicals, which are relatively easy to extract because they are on the surface of the leaf, have been widely used in such products as flavors and aromas, drugs, pesticides and cosmetics. Thus, the overexpression of genes that are used to produce diterpenes in plants may be accomplished by introducing transcription factor genes that induce said overexpression. One class of secondary metabolites, the diterpenes, can effect several biological systems such as tumor progression, prostaglandin synthesis and tissue inflammation. In addition, diterpenes can act as insect pheromones, termite allomones, and can exhibit neurotoxic, cytotoxic and antimutagenic activities. As a result of this functional diversity, diterpenes have been the target of research several pharmaceutical ventures. In most cases where the metabolic pathways are impossible to engineer, increasing trichome density or size on leaves may be the only way to increase plant productivity.

Production of anthocyanin in leaves and other plant parts. Several presently disclosed transcription factor genes can be used to alter anthocyanin production in numerous plant species. The potential utilities of these genes include alterations in pigment production for horticultural purposes, and possibly increasing stress resistance in combination with another transcription factor.

Production of miscellaneous secondary metabolites. Microarray data suggests that flux through the aromatic amino acid biosynthetic pathways and primary and secondary metabolite biosynthetic pathways are up-regulated. Presently disclosed transcription factors have been shown to be involved in regulating alkaloid biosynthesis, in part by up-regulating the enzymes indole-3-glycerol phosphatase and strictosidine synthase. Phenylalanine ammonia lyase, chalcone synthase and trans-cinnamate mono-oxygenase are also induced, and are involved in phenylpropanoid biosynthesis.

Sugar, starch, hemicellulose composition. Overexpression of the presently disclosed transcription factors that affect sugar content resulted in plants with altered leaf insoluble sugar content. Transcription factors that alter plant cell wall composition have several potential applications including altering food digestibility, plant tensile strength, wood quality, pathogen resistance and in pulp production. The

potential utilities of a gene involved in glucose-specific sugar sensing are to alter energy balance, photosynthetic rate, carbohydrate accumulation, biomass production, source-sink relationships, and senescence.

Hemicellulose is not desirable in paper pulps because of its lack of strength compared with cellulose. Thus modulating the amounts of cellulose vs. hemicellulose in the plant cell wall is desirable for the paper/lumber industry. Increasing the insoluble carbohydrate content in various fruits, vegetables, and other edible consumer products will result in enhanced fiber content. Increased fiber content would not only provide health benefits in food products, but might also increase digestibility of forage crops. In addition, the hemicellulose and pectin content of fruits and berries affects the quality of jam and catsup made from them. Changes in hemicellulose and pectin content could result in a superior consumer product.

Plant response to sugars and sugar composition. In addition to their important role as an energy source and structural component of the plant cell, sugars are central regulatory molecules that control several aspects of plant physiology, metabolism and development. It is thought that this control is achieved by regulating gene expression and, in higher plants, sugars have been shown to repress or activate plant genes involved in many essential processes such as photosynthesis, glyoxylate metabolism, respiration, starch and sucrose synthesis and degradation, pathogen response, wounding response, cell cycle regulation, pigmentation, flowering and senescence. The mechanisms by which sugars control gene expression are not understood.

Because sugars are important signaling molecules, the ability to control either the concentration of a signaling sugar or how the plant perceives or responds to a signaling sugar could be used to control plant development, physiology or metabolism. For example, the flux of sucrose (a disaccharide sugar used for systemically transporting carbon and energy in most plants) has been shown to affect gene expression and alter storage compound accumulation in seeds. Manipulation of the sucrose signaling pathway in seeds may therefore cause seeds to have more protein, oil or carbohydrate, depending on the type of manipulation. Similarly, in tubers, sucrose is converted to starch which is used as an energy store. It is thought that sugar signaling pathways may partially determine the levels of starch synthesized

in the tubers. The manipulation of sugar signaling in tubers could lead to tubers with a higher starch content.

Thus, the presently disclosed transcription factor genes that manipulate the sugar signal transduction pathway may lead to altered gene expression to produce plants with desirable traits. In particular, manipulation of sugar signal transduction pathways could be used to alter source-sink relationships in seeds, tubers, roots and other storage organs leading to increase in yield.

Plant growth rate and development. A number of the presently disclosed transcription factor genes have been shown to have significant effects on plant growth rate and development. These observations have included, for example, more rapid or delayed growth and development of reproductive organs. This would provide utility for regions with short or long growing seasons, respectively. Accelerating plant growth would also improve early yield or increase biomass at an earlier stage, when such is desirable (for example, in producing forestry products).

Embryo development. Presently disclosed transcription factor genes that alter embryo development has been used to alter seed protein and oil amounts and/or composition which is very important for the nutritional value and production of various food products. Seed shape and seed coat may also be altered by these genes, which may provide for improved storage stability.

Seed germination rate. A number of the presently disclosed transcription factor genes have been shown to modify seed germination rate, including when the seeds are in conditions normally unfavorable for germination (e.g., cold, heat or salt stress, or in the presence of ABA), and may thus be used to modify and improve germination rates under adverse conditions.

Plant, seedling vigor. Seedlings transformed with presently disclosed transcription factors have been shown to possess larger cotyledons and appeared somewhat more advanced than control plants. This indicates that the seedlings developed more rapidly than the control plants. Rapid seedling development is likely to reduce loss due to diseases particularly prevalent at the seedling stage (e.g.,

damping off) and is thus important for survivability of plants germinating in the field or in controlled environments.

Senescence, cell death. Presently disclosed transcription factor genes may be used to alter senescence responses in plants. Although leaf senescence is thought to be an evolutionary adaptation to recycle nutrients, the ability to control senescence in an agricultural setting has significant value. For example, a delay in leaf senescence in some maize hybrids is associated with a significant increase in yields and a delay of a few days in the senescence of soybean plants can have a large impact on yield. Delayed flower senescence may also generate plants that retain their blossoms longer and this may be of potential interest to the ornamental horticulture industry.

Modified fertility. Plants that overexpress a number of the presently disclosed transcription factor genes have been shown to possess reduced fertility. This could be a desirable trait, as it could be exploited to prevent or minimize the escape of the pollen of genetically modified organisms (GMOs) into the environment.

Early and delayed flowering. Presently disclosed transcription factor genes that accelerate flowering could have valuable applications in such programs since they allow much faster generation times. In a number of species, for example, broccoli, cauliflower, where the reproductive parts of the plants constitute the crop and the vegetative tissues are discarded, it would be advantageous to accelerate time to flowering. Accelerating flowering could shorten crop and tree breeding programs. Additionally, in some instances, a faster generation time might allow additional harvests of a crop to be made within a given growing season. A number of *Arabidopsis* genes have already been shown to accelerate flowering when constitutively expressed. These include LEAFY, APETALA1 and CONSTANS (Mandel, M. et al., 1995, Nature 377, 522-524; Weigel, D. and Nilsson, O., 1995, Nature 377, 495-500; Simon et al., 1996, Nature 384, 59-62).

By regulating the expression of potential flowering using inducible promoters, flowering could be triggered by application of an inducer chemical. This would allow flowering to be synchronized across a crop and facilitate more efficient harvesting. Such inducible systems could also be used to tune the flowering of crop varieties to

different latitudes. At present, species such as soybean and cotton are available as a series of maturity groups that are suitable for different latitudes on the basis of their flowering time (which is governed by day-length). A system in which flowering could be chemically controlled would allow a single high-yielding northern maturity group to be grown at any latitude. In southern regions such plants could be grown for longer, thereby increasing yields, before flowering was induced. In more northern areas, the induction would be used to ensure that the crop flowers prior to the first winter frosts.

In a sizeable number of species, for example, root crops, where the vegetative parts of the plants constitute the crop and the reproductive tissues are discarded, it would be advantageous to delay or prevent flowering. Extending vegetative development with presently disclosed transcription factor genes could thus bring about large increases in yields.. Prevention of flowering might help maximize vegetative yields and prevent escape of genetically modified organism (GMO) pollen.

Extended flowering phase. Presently disclosed transcription factors that extend flowering time have utility in engineering plants with longer-lasting flowers for the horticulture industry, and for extending the time in which the plant is fertile.

Flower and leaf development. Presently disclosed transcription factor genes have been used to modify the development of flowers and leaves. This could be advantageous in the development of new ornamental cultivars that present unique configurations. In addition, some of these genes have been shown to reduce a plant's fertility, which is also useful for helping to prevent development of pollen of GMOs.

Flower abscission. Presently disclosed transcription factor genes introduced into plants have been used to retain flowers for longer periods. This would provide a significant benefit to the ornamental industry, for both cut flowers and woody plant varieties (of, for example, maize), as well as have the potential to lengthen the fertile period of a plant, which could positively impact yield and breeding programs.

A listing of specific effects and utilities that the presently disclosed transcription factor genes have on plants, as determined by direct observation and assay analysis, is provided in Table 4.

XVI. Antisense and Co-suppression

In addition to expression of the nucleic acids of the invention as gene replacement or plant phenotype modification nucleic acids, the nucleic acids are also useful for sense and anti-sense suppression of expression, e.g., to down-regulate expression of a nucleic acid of the invention, e.g., as a further mechanism for modulating plant phenotype. That is, the nucleic acids of the invention, or subsequences or anti-sense sequences thereof, can be used to block expression of naturally occurring homologous nucleic acids. A variety of sense and anti-sense technologies are known in the art, e.g., as set forth in Lichtenstein and Nellen (1997) Antisense Technology: A Practical Approach IRL Press at Oxford University Press, Oxford, U.K.. In general, sense or anti-sense sequences are introduced into a cell, where they are optionally amplified, e.g., by transcription. Such sequences include both simple oligonucleotide sequences and catalytic sequences such as ribozymes.

For example, a reduction or elimination of expression (i.e., a "knock-out") of a transcription factor or transcription factor homologue polypeptide in a transgenic plant, e.g., to modify a plant trait, can be obtained by introducing an antisense construct corresponding to the polypeptide of interest as a cDNA. For antisense suppression, the transcription factor or homologue cDNA is arranged in reverse orientation (with respect to the coding sequence) relative to the promoter sequence in the expression vector. The introduced sequence need not be the full length cDNA or gene, and need not be identical to the cDNA or gene found in the plant type to be transformed. Typically, the antisense sequence need only be capable of hybridizing to the target gene or RNA of interest. Thus, where the introduced sequence is of shorter length, a higher degree of homology to the endogenous transcription factor sequence will be needed for effective antisense suppression. While antisense sequences of various lengths can be utilized, preferably, the introduced antisense sequence in the vector will be at least 30 nucleotides in length, and improved antisense suppression will typically be observed as the length of the antisense sequence increases. Preferably, the length of the antisense sequence in the vector will be greater than 100 nucleotides.

Transcription of an antisense construct as described results in the production of RNA molecules that are the reverse complement of mRNA molecules transcribed from the endogenous transcription factor gene in the plant cell.

Suppression of endogenous transcription factor gene expression can also be achieved using a ribozyme. Ribozymes are RNA molecules that possess highly specific endoribonuclease activity. The production and use of ribozymes are disclosed in U.S. Patent No. 4,987,071 and U.S. Patent No. 5,543,508. Synthetic ribozyme sequences including antisense RNAs can be used to confer RNA cleaving activity on the antisense RNA, such that endogenous mRNA molecules that hybridize to the antisense RNA are cleaved, which in turn leads to an enhanced antisense inhibition of endogenous gene expression.

Suppression of endogenous transcription factor gene expression can also be achieved using RNA interference, or RNAi. RNAi is a post-transcriptional, targeted gene-silencing technique that uses double-stranded RNA (dsRNA) to incite degradation of messenger RNA (mRNA) containing the same sequence as the dsRNA (Constans, (2002) *The Scientist* 16:36). Small interfering RNAs, or siRNAs are produced in at least two steps: an endogenous ribonuclease cleaves longer dsRNA into shorter, 21-23 nucleotide-long RNAs. The siRNA segments then mediate the degradation of the target mRNA (Zamore, (2001) *Nature Struct. Biol.*, 8:746-50). RNAi has been used for gene function determination in a manner similar to antisense oligonucleotides (Constans, (2002) *The Scientist* 16:36). Expression vectors that continually express siRNAs in transiently and stably transfected have been engineered to express small hairpin RNAs (shRNAs), which get processed in vivo into siRNAs-like molecules capable of carrying out gene-specific silencing (Brummelkamp et al., (2002) *Science* 296:550-553, and Paddison, et al. (2002) *Genes & Dev.* 16:948-958). Post-transcriptional gene silencing by double-stranded RNA is discussed in further detail by Hammond et al. (2001) *Nature Rev Gen* 2: 110-119, Fire et al. (1998) *Nature* 391: 806-811 and Timmons and Fire (1998) *Nature* 395: 854.

Vectors in which RNA encoded by a transcription factor or transcription factor homologue cDNA is over-expressed can also be used to obtain co-suppression of a corresponding endogenous gene, e.g., in the manner described in U.S. Patent No.

5,231,020 to Jorgensen. Such co-suppression (also termed sense suppression) does not require that the entire transcription factor cDNA be introduced into the plant cells, nor does it require that the introduced sequence be exactly identical to the endogenous transcription factor gene of interest. However, as with antisense suppression, the suppressive efficiency will be enhanced as specificity of hybridization is increased, e.g., as the introduced sequence is lengthened, and/or as the sequence similarity between the introduced sequence and the endogenous transcription factor gene is increased.

Vectors expressing an untranslatable form of the transcription factor mRNA, e.g., sequences comprising one or more stop codon, or nonsense mutation) can also be used to suppress expression of an endogenous transcription factor, thereby reducing or eliminating its activity and modifying one or more traits. Methods for producing such constructs are described in U.S. Patent No. 5,583,021. Preferably, such constructs are made by introducing a premature stop codon into the transcription factor gene. Alternatively, a plant trait can be modified by gene silencing using double-strand RNA (Sharp (1999) Genes and Development 13: 139-141). Another method for abolishing the expression of a gene is by insertion mutagenesis using the T-DNA of *Agrobacterium tumefaciens*. After generating the insertion mutants, the mutants can be screened to identify those containing the insertion in a transcription factor or transcription factor homologue gene. Plants containing a single transgene insertion event at the desired gene can be crossed to generate homozygous plants for the mutation. Such methods are well known to those of skill in the art. (See for example Koncz et al. (1992) Methods in Arabidopsis Research, World Scientific.)

Alternatively, a plant phenotype can be altered by eliminating an endogenous gene, such as a transcription factor or transcription factor homologue, e.g., by homologous recombination (Kempin et al. (1997) Nature 389:802-803).

A plant trait can also be modified by using the Cre-lox system (for example, as described in US Pat. No. 5,658,772). A plant genome can be modified to include first and second lox sites that are then contacted with a Cre recombinase. If the lox sites are in the same orientation, the intervening DNA sequence between the two sites

is excised. If the lox sites are in the opposite-orientation, the intervening sequence is inverted.

The polynucleotides and polypeptides of this invention can also be expressed in a plant in the absence of an expression cassette by manipulating the activity or expression level of the endogenous gene by other means. For example, by ectopically expressing a gene by T-DNA activation tagging (Ichikawa et al. (1997) Nature 390 698-701; Kakimoto et al. (1996) Science 274: 982-985). This method entails transforming a plant with a gene tag containing multiple transcriptional enhancers and once the tag has inserted into the genome, expression of a flanking gene coding sequence becomes deregulated. In another example, the transcriptional machinery in a plant can be modified so as to increase transcription levels of a polynucleotide of the invention (*See, e.g.,* PCT Publications WO 96/06166 and WO 98/53057 which describe the modification of the DNA-binding specificity of zinc finger proteins by changing particular amino acids in the DNA-binding motif).

The transgenic plant can also include the machinery necessary for expressing or altering the activity of a polypeptide encoded by an endogenous gene, for example by altering the phosphorylation state of the polypeptide to maintain it in an activated state.

Transgenic plants (or plant cells, or plant explants, or plant tissues) incorporating the polynucleotides of the invention and/or expressing the polypeptides of the invention can be produced by a variety of well established techniques as described above. Following construction of a vector, most typically an expression cassette, including a polynucleotide, e.g., encoding a transcription factor or transcription factor homologue, of the invention, standard techniques can be used to introduce the polynucleotide into a plant, a plant cell, a plant explant or a plant tissue of interest. Optionally, the plant cell, explant or tissue can be regenerated to produce a transgenic plant.

The plant can be any higher plant, including gymnosperms, monocotyledonous and dicotyledonous plants. Suitable protocols are available for *Leguminosae* (alfalfa, soybean, clover, etc.), *Umbelliferae* (carrot, celery, parsnip),

Cruciferae (cabbage, radish, rapeseed, broccoli, etc.), *Curcubitaceae* (melons and cucumber), *Gramineae* (wheat, corn, rice, barley, millet, etc.), *Solanaceae* (potato, tomato, tobacco, peppers, etc.), and various other crops. See protocols described in Ammirato et al. (1984) Handbook of Plant Cell Culture –Crop Species, Macmillan Publ. Co. Shimamoto et al. (1989) Nature 338:274-276; Fromm et al. (1990) Bio/Technology 8:833-839; and Vasil et al. (1990) Bio/Technology 8:429-434.

Transformation and regeneration of both monocotyledonous and dicotyledonous plant cells is now routine, and the selection of the most appropriate transformation technique will be determined by the practitioner. The choice of method will vary with the type of plant to be transformed; those skilled in the art will recognize the suitability of particular methods for given plant types. Suitable methods can include, but are not limited to: electroporation of plant protoplasts; liposome-mediated transformation; polyethylene glycol (PEG) mediated transformation; transformation using viruses; micro-injection of plant cells; micro-projectile bombardment of plant cells; vacuum infiltration; and *Agrobacterium tumefaciens* mediated transformation. Transformation means introducing a nucleotide sequence into a plant in a manner to cause stable or transient expression of the sequence.

Successful examples of the modification of plant characteristics by transformation with cloned sequences which serve to illustrate the current knowledge in this field of technology, and which are herein incorporated by reference, include: U.S. Patent Nos. 5,571,706; 5,677,175; 5,510,471; 5,750,386; 5,597,945; 5,589,615; 5,750,871; 5,268,526; 5,780,708; 5,538,880; 5,773,269; 5,736,369 and 5,610,042.

Following transformation, plants are preferably selected using a dominant selectable marker incorporated into the transformation vector. Typically, such a marker will confer antibiotic or herbicide resistance on the transformed plants, and selection of transformants can be accomplished by exposing the plants to appropriate concentrations of the antibiotic or herbicide.

After transformed plants are selected and grown to maturity, those plants showing a modified trait are identified. The modified trait can be any of those traits described above. Additionally, to confirm that the modified trait is due to changes in

expression levels or activity of the polypeptide or polynucleotide of the invention can be determined by analyzing mRNA expression using Northern blots, RT-PCR or microarrays, or protein expression using immunoblots or Western blots or gel shift assays.

XVII. Integrated Systems – Sequence Identity

Additionally, the present invention may be an integrated system, computer or computer readable medium that comprises an instruction set for determining the identity of one or more sequences in a database. In addition, the instruction set can be used to generate or identify sequences that meet any specified criteria. Furthermore, the instruction set may be used to associate or link certain functional benefits, such improved characteristics, with one or more identified sequence.

For example, the instruction set can include, e.g., a sequence comparison or other alignment program, e.g., an available program such as, for example, the Wisconsin Package Version 10.0, such as BLAST, FASTA, PILEUP, FINDPATTERNS or the like (GCG, Madison, WI). Public sequence databases such as GenBank, EMBL, Swiss-Prot and PIR or private sequence databases such as PHYTOSEQ sequence database (Incyte Genomics, Palo Alto, CA) can be searched.

Alignment of sequences for comparison can be conducted by the local homology algorithm of Smith and Waterman (1981) Adv. Appl. Math. 2:482, by the homology alignment algorithm of Needleman and Wunsch (1970) J. Mol. Biol. 48:443-453, by the search for similarity method of Pearson and Lipman (1988) Proc. Natl. Acad. Sci. U.S.A. 85:2444-2448, by computerized implementations of these algorithms. After alignment, sequence comparisons between two (or more) polynucleotides or polypeptides are typically performed by comparing sequences of the two sequences over a comparison window to identify and compare local regions of sequence similarity. The comparison window can be a segment of at least about 20 contiguous positions, usually about 50 to about 200, more usually about 100 to about 150 contiguous positions. A description of the method is provided in Ausubel et al., *supra*.

A variety of methods for determining sequence relationships can be used, including manual alignment and computer assisted sequence alignment and analysis. This later approach is a preferred approach in the present invention, due to the increased throughput afforded by computer assisted methods. As noted above, a variety of computer programs for performing sequence alignment are available, or can be produced by one of skill.

One example algorithm that is suitable for determining percent sequence identity and sequence similarity is the BLAST algorithm, which is described in Altschul et al. J. Mol. Biol. 215:403-410 (1990). Software for performing BLAST analyses is publicly available, e.g., through the National Center for Biotechnology Information (see internet website at ncbi.nlm.nih.gov). This algorithm involves first identifying high scoring sequence pairs (HSPs) by identifying short words of length W in the query sequence, which either match or satisfy some positive-valued threshold score T when aligned with a word of the same length in a database sequence. T is referred to as the neighborhood word score threshold (Altschul et al., *supra*). These initial neighborhood word hits act as seeds for initiating searches to find longer HSPs containing them. The word hits are then extended in both directions along each sequence for as far as the cumulative alignment score can be increased. Cumulative scores are calculated using, for nucleotide sequences, the parameters M (reward score for a pair of matching residues; always > 0) and N (penalty score for mismatching residues; always < 0). For amino acid sequences, a scoring matrix is used to calculate the cumulative score. Extension of the word hits in each direction are halted when: the cumulative alignment score falls off by the quantity X from its maximum achieved value; the cumulative score goes to zero or below, due to the accumulation of one or more negative-scoring residue alignments; or the end of either sequence is reached. The BLAST algorithm parameters W, T, and X determine the sensitivity and speed of the alignment. The BLASTN program (for nucleotide sequences) uses as defaults a wordlength (W) of 11, an expectation (E) of 10, a cutoff of 100, M=5, N=-4, and a comparison of both strands. For amino acid sequences, the BLASTP program uses as defaults a wordlength (W) of 3, an expectation (E) of 10, and the BLOSUM62 scoring matrix (*see* Henikoff & Henikoff (1989) Proc. Natl. Acad. Sci. USA 89:10915). Unless otherwise indicated, "sequence identity" here refers to the % sequence identity generated from a tblastx using the NCBI version of

the algorithm at the default settings using gapped alignments with the filter "off" (see, for example, internet website at ncbi.nlm.nih.gov).

In addition to calculating percent sequence identity, the BLAST algorithm also performs a statistical analysis of the similarity between two sequences (*see, e.g.,* Karlin & Altschul (1993) Proc. Natl. Acad. Sci. USA 90:5873-5787). One measure of similarity provided by the BLAST algorithm is the smallest sum probability ($P(N)$), which provides an indication of the probability by which a match between two nucleotide or amino acid sequences would occur by chance. For example, a nucleic acid is considered similar to a reference sequence (and, therefore, in this context, homologous) if the smallest sum probability in a comparison of the test nucleic acid to the reference nucleic acid is less than about 0.1, or less than about 0.01, and or even less than about 0.001. An additional example of a useful sequence alignment algorithm is PILEUP. PILEUP creates a multiple sequence alignment from a group of related sequences using progressive, pairwise alignments. The program can align, e.g., up to 300 sequences of a maximum length of 5,000 letters.

The integrated system, or computer typically includes a user input interface allowing a user to selectively view one or more sequence records corresponding to the one or more character strings, as well as an instruction set which aligns the one or more character strings with each other or with an additional character string to identify one or more region of sequence similarity. The system may include a link of one or more character strings with a particular phenotype or gene function. Typically, the system includes a user readable output element that displays an alignment produced by the alignment instruction set.

The methods of this invention can be implemented in a localized or distributed computing environment. In a distributed environment, the methods may implemented on a single computer comprising multiple processors or on a multiplicity of computers. The computers can be linked, e.g. through a common bus, but more preferably the computer(s) are nodes on a network. The network can be a generalized or a dedicated local or wide-area network and, in certain preferred embodiments, the computers may be components of an intra-net or an internet.

Thus, the invention provides methods for identifying a sequence similar or homologous to one or more polynucleotides as noted herein, or one or more target polypeptides encoded by the polynucleotides, or otherwise noted herein and may include linking or associating a given plant phenotype or gene function with a sequence. In the methods, a sequence database is provided (locally or across an inter or intra net) and a query is made against the sequence database using the relevant sequences herein and associated plant phenotypes or gene functions.

Any sequence herein can be entered into the database, before or after querying the database. This provides for both expansion of the database and, if done before the querying step, for insertion of control sequences into the database. The control sequences can be detected by the query to ensure the general integrity of both the database and the query. As noted, the query can be performed using a web browser based interface. For example, the database can be a centralized public database such as those noted herein, and the querying can be done from a remote terminal or computer across an internet or intranet.

XVIII. Examples

The following examples are intended to illustrate but not limit the present invention. The complete descriptions of the traits associated with each polynucleotide of the invention is fully disclosed in Table 4 and Table 6.

Example I: Full Length Gene Identification and Cloning

Putative transcription factor sequences (genomic or ESTs) related to known transcription factors were identified in the *Arabidopsis thaliana* GenBank database using the tblastn sequence analysis program using default parameters and a P-value cutoff threshold of -4 or -5 or lower, depending on the length of the query sequence. Putative transcription factor sequence hits were then screened to identify those containing particular sequence strings. If the sequence hits contained such sequence strings, the sequences were confirmed as transcription factors.

Alternatively, *Arabidopsis thaliana* cDNA libraries derived from different tissues or treatments, or genomic libraries were screened to identify novel members of a transcription family using a low stringency hybridization approach. Probes were

synthesized using gene specific primers in a standard PCR reaction (annealing temperature 60° C) and labeled with ³²P dCTP using the High Prime DNA Labeling Kit (Boehringer Mannheim). Purified radiolabelled probes were added to filters immersed in Church hybridization medium (0.5 M NaPO₄ pH 7.0, 7% SDS, 1 % w/v bovine serum albumin) and hybridized overnight at 60°C with shaking. Filters were washed two times for 45 to 60 minutes with 1xSSC, 1% SDS at 60° C.

To identify additional sequence 5' or 3' of a partial cDNA sequence in a cDNA library, 5' and 3' rapid amplification of cDNA ends (RACE) was performed using the U.C. Marathon cDNA amplification kit (Clontech, Palo Alto, CA). Generally, the method entailed first isolating poly(A) mRNA, performing first and second strand cDNA synthesis to generate double stranded cDNA, blunting cDNA ends, followed by ligation of the U.C. Marathon Adaptor to the cDNA to form a library of adaptor-ligated ds cDNA.

Gene-specific primers were designed to be used along with adaptor specific primers for both 5' and 3' RACE reactions. Nested primers, rather than single primers, were used to increase PCR specificity. Using 5' and 3' RACE reactions, 5' and 3' RACE fragments were obtained, sequenced and cloned. The process can be repeated until 5' and 3' ends of the full-length gene were identified. Then the full-length cDNA was generated by PCR using primers specific to 5' and 3' ends of the gene by end-to-end PCR.

Example II: Construction of Expression Vectors

The sequence was amplified from a genomic or cDNA library using primers specific to sequences upstream and downstream of the coding region. The expression vector was pMEN20 or pMEN65, which are both derived from pMON316 (Sanders et al, (1987) Nucleic Acids Research 15:1543-1558) and contain the CaMV 35S promoter to express transgenes. To clone the sequence into the vector, both pMEN20 and the amplified DNA fragment were digested separately with SalI and NotI restriction enzymes at 37° C for 2 hours. The digestion products were subject to electrophoresis in a 0.8% agarose gel and visualized by ethidium bromide staining. The DNA fragments containing the sequence and the linearized plasmid were excised and purified by using a Qiaquick gel extraction kit (Qiagen, Valencia CA). The

fragments of interest were ligated at a ratio of 3:1 (vector to insert). Ligation reactions using T4 DNA ligase (New England Biolabs, Beverly MA) were carried out at 16° C for 16 hours. The ligated DNAs were transformed into competent cells of the *E. coli* strain DH5alpha by using the heat shock method. The transformations were plated on LB plates containing 50 mg/l kanamycin (Sigma, St. Louis, MO). Individual colonies were grown overnight in five milliliters of LB broth containing 50 mg/l kanamycin at 37° C. Plasmid DNA was purified by using Qiaquick Mini Prep kits (Qiagen).

Example III: Transformation of *Agrobacterium* with the Expression Vector

After the plasmid vector containing the gene was constructed, the vector was used to transform *Agrobacterium tumefaciens* cells expressing the gene products. The stock of *Agrobacterium tumefaciens* cells for transformation were made as described by Nagel et al. (1990) FEMS Microbiol Letts. 67: 325-328. *Agrobacterium* strain ABI was grown in 250 ml LB medium (Sigma) overnight at 28°C with shaking until an absorbance (A_{600}) of 0.5 – 1.0 was reached. Cells were harvested by centrifugation at 4,000 x g for 15 min at 4° C. Cells were then resuspended in 250 µl chilled buffer (1 mM HEPES, pH adjusted to 7.0 with KOH). Cells were centrifuged again as described above and resuspended in 125 µl chilled buffer. Cells were then centrifuged and resuspended two more times in the same HEPES buffer as described above at a volume of 100 µl and 750 µl, respectively. Resuspended cells were then distributed into 40 µl aliquots, quickly frozen in liquid nitrogen, and stored at -80° C.

Agrobacterium cells were transformed with plasmids prepared as described above following the protocol described by Nagel et al. For each DNA construct to be transformed, 50 – 100 ng DNA (generally resuspended in 10 mM Tris-HCl, 1 mM EDTA, pH 8.0) was mixed with 40 µl of *Agrobacterium* cells. The DNA/cell mixture was then transferred to a chilled cuvette with a 2mm electrode gap and subject to a 2.5 kV charge dissipated at 25 µF and 200 µF using a Gene Pulser II apparatus (Bio-Rad, Hercules, CA). After electroporation, cells were immediately resuspended in 1.0 ml LB and allowed to recover without antibiotic selection for 2 – 4 hours at 28° C in a shaking incubator. After recovery, cells were plated onto selective medium of LB broth containing 100 µg/ml spectinomycin (Sigma) and incubated for 24-48 hours at 28° C. Single colonies were then picked and inoculated in fresh medium. The

presence of the plasmid construct was verified by PCR amplification and sequence analysis.

Example IV: Transformation of *Arabidopsis* Plants with *Agrobacterium tumefaciens* with Expression Vector

After transformation of *Agrobacterium tumefaciens* with plasmid vectors containing the gene, single *Agrobacterium* colonies were identified, propagated, and used to transform *Arabidopsis* plants. Briefly, 500 ml cultures of LB medium containing 50 mg/l kanamycin were inoculated with the colonies and grown at 28° C with shaking for 2 days until an optical absorbance at 600 nm wavelength over 1 cm (A_{600}) of > 2.0 is reached. Cells were then harvested by centrifugation at 4,000 x g for 10 min, and resuspended in infiltration medium (1/2 X Murashige and Skoog salts (Sigma), 1 X Gamborg's B-5 vitamins (Sigma), 5.0% (w/v) sucrose (Sigma), 0.044 μ M benzylamino purine (Sigma), 200 μ l/l Silwet L-77 (Lehle Seeds) until an A_{600} of 0.8 was reached.

Prior to transformation, *Arabidopsis thaliana* seeds (ecotype Columbia) were sown at a density of ~10 plants per 4" pot onto Pro-Mix BX potting medium (Hummert International) covered with fiberglass mesh (18 mm X 16 mm). Plants were grown under continuous illumination (50-75 μ E/m²/sec) at 22-23° C with 65-70% relative humidity. After about 4 weeks, primary inflorescence stems (bolts) are cut off to encourage growth of multiple secondary bolts. After flowering of the mature secondary bolts, plants were prepared for transformation by removal of all siliques and opened flowers.

The pots were then immersed upside down in the mixture of *Agrobacterium* infiltration medium as described above for 30 sec, and placed on their sides to allow draining into a 1' x 2' flat surface covered with plastic wrap. After 24 h, the plastic wrap was removed and pots are turned upright. The immersion procedure was repeated one week later, for a total of two immersions per pot. Seeds were then collected from each transformation pot and analyzed following the protocol described below.

Example V: Identification of *Arabidopsis* Primary Transformants

Seeds collected from the transformation pots were sterilized essentially as follows. Seeds were dispersed into in a solution containing 0.1% (v/v) Triton X-100 (Sigma) and sterile H₂O and washed by shaking the suspension for 20 min. The wash solution was then drained and replaced with fresh wash solution to wash the seeds for 20 min with shaking. After removal of the second wash solution, a solution containing 0.1% (v/v) Triton X-100 and 70% ethanol (Equistar) was added to the seeds and the suspension was shaken for 5 min. After removal of the ethanol/detergent solution, a solution containing 0.1% (v/v) Triton X-100 and 30% (v/v) bleach (Clorox) was added to the seeds, and the suspension was shaken for 10 min. After removal of the bleach/detergent solution, seeds were then washed five times in sterile distilled H₂O. The seeds were stored in the last wash water at 4° C for 2 days in the dark before being plated onto antibiotic selection medium (1 X Murashige and Skoog salts (pH adjusted to 5.7 with 1M KOH), 1 X Gamborg's B-5 vitamins, 0.9% phytagar (Life Technologies), and 50 mg/l kanamycin). Seeds were germinated under continuous illumination (50-75 $\mu\text{E}/\text{m}^2/\text{sec}$) at 22-23° C. After 7-10 days of growth under these conditions, kanamycin resistant primary transformants (T₁ generation) were visible and obtained. These seedlings were transferred first to fresh selection plates where the seedlings continued to grow for 3-5 more days, and then to soil (Pro-Mix BX potting medium).

Primary transformants were crossed and progeny seeds (T₂) collected; kanamycin resistant seedlings were selected and analyzed. The expression levels of the recombinant polynucleotides in the transformants varies from about a 5% expression level increase to a least a 100% expression level increase. Similar observations are made with respect to polypeptide level expression.

Example VI: Identification of *Arabidopsis* Plants with Transcription Factor Gene Knockouts

The screening of insertion mutagenized *Arabidopsis* collections for null mutants in a known target gene was essentially as described in Krysan et al (1999) Plant Cell 11:2283-2290. Briefly, gene-specific primers, nested by 5-250 base pairs to each other, were designed from the 5' and 3' regions of a known target gene. Similarly, nested sets of primers were also created specific to each of the T-DNA or

transposon ends (the "right" and "left" borders). All possible combinations of gene specific and T-DNA/transposon primers were used to detect by PCR an insertion event within or close to the target gene. The amplified DNA fragments were then sequenced which allows the precise determination of the T-DNA/transposon insertion point relative to the target gene. Insertion events within the coding or intervening sequence of the genes were deconvoluted from a pool comprising a plurality of insertion events to a single unique mutant plant for functional characterization. The method is described in more detail in Yu and Adam, US Application Serial No. 09/177,733 filed October 23, 1998.

Example VII: Identification of Modified Phenotypes in Overexpression or Gene Knockout Plants

Experiments were performed to identify those transformants or knockouts that exhibited modified biochemical characteristics. Among the biochemicals that were assayed were insoluble sugars, such as arabinose, fucose, galactose, mannose, rhamnose or xylose or the like; prenyl lipids, such as lutein, beta-carotene, xanthophyll-1, xanthophyll-2, chlorophylls A or B, or alpha-, delta- or gamma-tocopherol or the like; fatty acids, such as 16:0 (palmitic acid), 16:1 (palmitoleic acid), 18:0 (stearic acid), 18:1 (oleic acid), 18:2 (linoleic acid), 20:0, 18:3 (linolenic acid), 20:1 (eicosenoic acid), 20:2, 22:1 (erucic acid) or the like; waxes, such as by altering the levels of C29, C31, or C33 alkanes; sterols, such as brassicasterol, campesterol, stigmasterol, sitosterol or stigmastanol or the like, glucosinolates, protein or oil levels.

Fatty acids were measured using two methods depending on whether the tissue was from leaves or seeds. For leaves, lipids were extracted and esterified with hot methanolic H_2SO_4 and partitioned into hexane from methanolic brine. For seed fatty acids, seeds were pulverized and extracted in methanol:heptane:toluene:2,2-dimethoxypropane: H_2SO_4 (39:34:20:5:2) for 90 minutes at 80°C. After cooling to room temperature the upper phase, containing the seed fatty acid esters, was subjected to GC analysis. Fatty acid esters from both seed and leaf tissues were analyzed with a Supelco SP-2330 column.

Glucosinolates were purified from seeds or leaves by first heating the tissue at 95°C for 10 minutes. Preheated ethanol:water (50:50) is and after heating at 95°C for a further 10 minutes, the extraction solvent is applied to a DEAE Sephadex column which had been previously equilibrated with 0.5 M pyridine acetate. Desulfoglucosinolates were eluted with 300 ul water and analyzed by reverse phase HPLC monitoring at 226 nm.

For wax alkanes, samples were extracted using an identical method as fatty acids and extracts were analyzed on a HP 5890 GC coupled with a 5973 MSD. Samples were chromatographically isolated on a J&W DB35 mass spectrometer (J&W Scientific).

To measure prenyl lipids levels, seeds or leaves were pulverized with 1 to 2% pyrogallol as an antioxidant. For seeds, extracted samples were filtered and a portion removed for tocopherol and carotenoid/chlorophyll analysis by HPLC. The remaining material was saponified for sterol determination. For leaves, an aliquot was removed and diluted with methanol and chlorophyll A, chlorophyll B, and total carotenoids measured by spectrophotometry by determining optical absorbance at 665.2 nm, 652.5 nm, and 470 nm. An aliquot was removed for tocopherol and carotenoid/chlorophyll composition by HPLC using a Waters uBondapak C18 column (4.6 mm x 150 mm). The remaining methanolic solution was saponified with 10% KOH at 80°C for one hour. The samples were cooled and diluted with a mixture of methanol and water. A solution of 2% methylene chloride in hexane was mixed in and the samples were centrifuged. The aqueous methanol phase was again re-extracted 2% methylene chloride in hexane and, after centrifugation, the two upper phases were combined and evaporated. 2% methylene chloride in hexane was added to the tubes and the samples were then extracted with one ml of water. The upper phase was removed, dried, and resuspended in 400 ul of 2% methylene chloride in hexane and analyzed by gas chromatography using a 50 m DB-5ms (0.25 mm ID, 0.25 um phase, J&W Scientific).

Insoluble sugar levels were measured by the method essentially described by Reiter et al., (1997) Plant Journal 12:335-345. This method analyzes the neutral sugar composition of cell wall polymers found in *Arabidopsis* leaves. Soluble sugars were

separated from sugar polymers by extracting leaves with hot 70% ethanol. The remaining residue containing the insoluble polysaccharides was then acid hydrolyzed with allose added as an internal standard. Sugar monomers generated by the hydrolysis were then reduced to the corresponding alditols by treatment with NaBH_4 , then were acetylated to generate the volatile alditol acetates which were then analyzed by GC-FID. Identity of the peaks was determined by comparing the retention times of known sugars converted to the corresponding alditol acetates with the retention times of peaks from wild-type plant extracts. Alditol acetates were analyzed on a Supelco SP-2330 capillary column (30 m x 250 μm x 0.2 μm) using a temperature program beginning at 180° C for 2 minutes followed by an increase to 220° C in 4 minutes. After holding at 220° C for 10 minutes, the oven temperature is increased to 240° C in 2 minutes and held at this temperature for 10 minutes and brought back to room temperature.

To identify plants with alterations in total seed oil or protein content, 150mg of seeds from T2 progeny plants were subjected to analysis by Near Infrared Reflectance Spectroscopy (NIRS) using a Foss NirSystems Model 6500 with a spinning cup transport system. NIRS is a non-destructive analytical method used to determine seed oil and protein composition. Infrared is the region of the electromagnetic spectrum located after the visible region in the direction of longer wavelengths. 'Near infrared' owns its name for being the infrared region near to the visible region of the electromagnetic spectrum. For practical purposes, near infrared comprises wavelengths between 800 and 2500 nm. NIRS is applied to organic compounds rich in O-H bonds (such as moisture, carbohydrates, and fats), C-H bonds (such as organic compounds and petroleum derivatives), and N-H bonds (such as proteins and amino acids). The NIRS analytical instruments operate by statistically correlating NIRS signals at several wavelengths with the characteristic or property intended to be measured. All biological substances contain thousands of C-H, O-H, and N-H bonds. Therefore, the exposure to near infrared radiation of a biological sample, such as a seed, results in a complex spectrum which contains qualitative and quantitative information about the physical and chemical composition of that sample.

The numerical value of a specific analyte in the sample, such as protein content or oil content, is mediated by a calibration approach known as chemometrics.

Chemometrics applies statistical methods such as multiple linear regression (MLR), partial least squares (PLS), and principle component analysis (PCA) to the spectral data and correlates them with a physical property or other factor, that property or factor is directly determined rather than the analyte concentration itself. The method first provides "wet chemistry" data of the samples required to develop the calibration.

Calibration for Arabidopsis seed oil composition was performed using accelerated solvent extraction using 1 g seed sample size and was validated against certified canola seed. A similar wet chemistry approach was performed for seed protein composition calibration.

Data obtained from NIRS analysis was analyzed statistically using a nearest-neighbor (N-N) analysis. The N-N analysis allows removal of within-block spatial variability in a fairly flexible fashion which does not require prior knowledge of the pattern of variability in the chamber. Ideally, all hybrids are grown under identical experimental conditions within a block (rep). In reality, even in many block designs, significant within-block variability exists. Nearest-neighbor procedures are based on assumption that environmental effect of a plot is closely related to that of its neighbors. Nearest-neighbor methods use information from adjacent plots to adjust for within-block heterogeneity and so provide more precise estimates of treatment means and differences. If there is within-plot heterogeneity on a spatial scale that is larger than a single plot and smaller than the entire block, then yields from adjacent plots will be positively correlated. Information from neighboring plots can be used to reduce or remove the unwanted effect of the spatial heterogeneity, and hence improve the estimate of the treatment effect. Data from neighboring plots can also be used to reduce the influence of competition between adjacent plots. The Papadakis N-N analysis can be used with designs to remove within-block variability that would not be removed with the standard split plot analysis (Papadakis, 1973, Inst. d'Amelior. Plantes Thessaloniki (Greece) Bull. Scientif., No. 23; Papadakis, 1984, Proc. Acad. Athens, 59, 326-342).

Experiments were performed to identify those transformants or knockouts that exhibited an improved pathogen tolerance. For such studies, the transformants were exposed to biotrophic fungal pathogens, such as *Erysiphe orontii*, and necrotropic

fungus pathogens, such as *Fusarium oxysporum*. *Fusarium oxysporum* isolates cause vascular wilts and damping off of various annual vegetables, perennials and weeds (Mauch-Mani and Slusarenko (1994) Molecular Plant-Microbe Interactions 7: 378-383). For *Fusarium oxysporum* experiments, plants grown on Petri dishes were sprayed with a fresh spore suspension of *F. oxysporum*. The spore suspension was prepared as follows: A plug of fungal hyphae from a plate culture was placed on a fresh potato dextrose agar plate and allowed to spread for one week. 5 ml sterile water was then added to the plate, swirled, and pipetted into 50 ml Armstrong *Fusarium* medium. Spores were grown overnight in *Fusarium* medium and then sprayed onto plants using a Preval paint sprayer. Plant tissue was harvested and frozen in liquid nitrogen 48 hours post infection.

Erysiphe orontii is a causal agent of powdery mildew. For *Erysiphe orontii* experiments, plants were grown approximately 4 weeks in a greenhouse under 12 hour light (20°C, ~30% relative humidity (rh)). Individual leaves were infected with *E. orontii* spores from infected plants using a camel's hair brush, and the plants were transferred to a Percival growth chamber (20°C, 80% rh.). Plant tissue was harvested and frozen in liquid nitrogen 7 days post infection.

Botrytis cinerea is a necrotrophic pathogen. *Botrytis cinerea* was grown on potato dextrose agar in the light. A spore culture was made by spreading 10 ml of sterile water on the fungus plate, swirling and transferring spores to 10 ml of sterile water. The spore inoculum (approx. 10⁵ spores/ml) was used to spray 10 day-old seedlings grown under sterile conditions on MS (minus sucrose) media. Symptoms were evaluated every day up to approximately 1 week.

Infection with bacterial pathogens *Pseudomonas syringae* pv *maculicola* (Psm) strain 4326 and pv *maculicola* strain 4326 was performed by hand inoculation at two doses. Two inoculation doses allows the differentiation between plants with enhanced susceptibility and plants with enhanced resistance to the pathogen. Plants were grown for 3 weeks in the greenhouse, then transferred to the growth chamber for the remainder of their growth. Psm ES4326 was hand inoculated with 1 ml syringe on 3 fully-expanded leaves per plant (4 1/2 wk old), using at least 9 plants per overexpressing line at two inoculation doses, OD=0.005 and OD=0.0005. Disease

scoring occurred at day 3 post-inoculation with pictures of the plants and leaves taken in parallel.

In some instances, expression patterns of the pathogen-induced genes (such as defense genes) was monitored by microarray experiments. cDNAs were generated by PCR and resuspended at a final concentration of ~ 100 ng/ul in 3X SSC or 150mM Na-phosphate (Eisen and Brown (1999) *Methods Enzymol.* 303:179-205). The cDNAs were spotted on microscope glass slides coated with polylysine. The prepared cDNAs were aliquoted into 384 well plates and spotted on the slides using an x-y-z gantry (OmniGrid) purchased from GeneMachines (Menlo Park, CA) outfitted with quill type pins purchased from Telechem International (Sunnyvale, CA). After spotting, the arrays were cured for a minimum of one week at room temperature, rehydrated and blocked following the protocol recommended by Eisen and Brown (1999; *supra*).

Sample total RNA (10 ug) samples were labeled using fluorescent Cy3 and Cy5 dyes. Labeled samples were resuspended in 4X SSC/0.03% SDS/4 ug salmon sperm DNA/2 ug tRNA/ 50mM Na-pyrophosphate, heated for 95°C for 2.5 minutes, spun down and placed on the array. The array was then covered with a glass coverslip and placed in a sealed chamber. The chamber was then kept in a water bath at 62°C overnight. The arrays were washed as described in Eisen and Brown (1999) and scanned on a General Scanning 3000 laser scanner. The resulting files are subsequently quantified using Image, a software purchased from BioDiscovery (Los Angeles, CA).

Experiments were performed to identify those transformants or knockouts that exhibited an improved environmental stress tolerance. For such studies, the transformants were exposed to a variety of environmental stresses. Plants were exposed to chilling stress (6 hour exposure to 4-8° C), heat stress (6 hour exposure to 32-37° C), high salt stress (6 hour exposure to 200 mM NaCl), drought stress (168 hours after removing water from trays), osmotic stress (6 hour exposure to 3 M mannitol), or nutrient limitation (nitrogen, phosphate, and potassium) (Nitrogen: all components of MS medium remained constant except N was reduced to 20 mg/l of NH₄NO₃, or Phosphate: All components of MS medium except KH₂PO₄, which was

replaced by K_2SO_4 , Potassium: All components of MS medium except removal of KNO_3 and KH_2PO_4 , which were replaced by NaH_4PO_4).

Experiments were performed to identify those transformants or knockouts that exhibited a modified structure and development characteristics. For such studies, the transformants were observed by eye to identify novel structural or developmental characteristics associated with the ectopic expression of the polynucleotides or polypeptides of the invention.

Experiments were performed to identify those transformants or knockouts that exhibited modified sugar-sensing. For such studies, seeds from transformants were germinated on media containing 5% glucose or 9.4% sucrose which normally partially restrict hypocotyl elongation. Plants with altered sugar sensing may have either longer or shorter hypocotyls than normal plants when grown on this media. Additionally, other plant traits may be varied such as root mass.

Flowering time was measured by the number of rosette leaves present when a visible inflorescence of approximately 3 cm is apparent. Rosette and total leaf number on the progeny stem are tightly correlated with the timing of flowering (Koorneef et al (1991) *Mol. Gen. Genet* 229:57-66). The vernalization response was measured. For vernalization treatments, seeds were sown to MS agar plates, sealed with micropore tape, and placed in a 4°C cold room with low light levels for 6-8 weeks. The plates were then transferred to the growth rooms alongside plates containing freshly sown non-vernalized controls. Rosette leaves were counted when a visible inflorescence of approximately 3 cm was apparent.

Modified phenotypes observed for particular overexpressor or knockout plants are provided in Table 4. For a particular overexpressor that shows a less beneficial characteristic, it may be more useful to select a plant with a decreased expression of the particular transcription factor. For a particular knockout that shows a less beneficial characteristic, it may be more useful to select a plant with an increased expression of the particular transcription factor.

The sequences of the Sequence Listing or those in Tables 4 , 5 or those disclosed here can be used to prepare transgenic plants and plants with altered traits. The specific transgenic plants listed below are produced from the sequences of the Sequence Listing, as noted. Table 4 provides exemplary polynucleotide and polypeptide sequences of the invention. Table 4 includes, from left to right for each sequence: the first column shows the polynucleotide SEQ ID NO; the second column shows the Mendel Gene ID No., GID; the third column shows the trait(s) resulting from the knock out or overexpression of the polynucleotide in the transgenic plant; the fourth column shows the category of the trait; the fifth column shows the transcription factor family to which the polynucleotide belongs; the sixth column ("Comment"), includes specific effects and utilities conferred by the polynucleotide of the first column; the seventh column shows the SEQ ID NO of the polypeptide encoded by the polynucleotide; and the eighth column shows the amino acid residue positions of the conserved domain in amino acid (AA) co-ordinates.

Seed of plants overexpressing sequences G265 (SEQ ID NOs:871 and 872), G715 (SEQ ID NOs:925 and 926), G1471 (SEQ ID NOs:311 and 312), G1793 (SEQ ID NOs:365 and 366), G1838 (SEQ ID NOs:381 and 382), G1902 (SEQ ID NOs:405 and 406), G286 (SEQ ID NOs:877 and 878), G2138 (SEQ ID NOs:865 and 866) and G2830 (SEQ ID NOs:875 and 876) was subjected to NIR analysis and a significant increase in seed oil content compared with seed from control plants was identified.

G192: G192 (SEQ ID NO: 859) was expressed in all plant tissues and under all conditions examined. Its expression was slightly induced upon infection by *Fusarium*. G192 was analyzed using transgenic plants in which this gene was expressed under the control of the 35S promoter. G192 overexpressors were late flowering under 12 hour light and had more leaves than control plants. This phenotype was manifested in the three T2 lines analyzed. Results of one experiment suggest that G192 overexpressor was more susceptible to infection with a moderate dose of the fungal pathogen *Erysiphe orontii*. The decrease in seed oil observed for one line was replicated in an independent experiment. G192 overexpression delayed flowering. A wide variety of applications exist for systems that either lengthen or shorten the time to flowering, or for systems of inducible flowering time control. In particular, in species where the vegetative parts of the plants constitute the crop and

the reproductive tissues are discarded, it will be advantageous to delay or prevent flowering. Extending vegetative development can bring about large increases in yields. G192 can be used to manipulate the defense response in order to generate pathogen-resistant plants. G192 can be used to manipulate seed oil content, which can be of nutritional value.

Closely Related Genes from Other Species

G192 had some similarity within the conserved WRKY domain to non-Arabidopsis plant proteins.

G1946: G1946 (SEQ ID NO: 801) was studied using transgenic plants in which the gene was expressed under the control of the 35S promoter. Overexpression of G1946 resulted in accelerated flowering, with 35S::G1946 transformants producing flower buds up to a week earlier than wild-type controls (24-hour light conditions). These effects were seen in 12/20 primary transformants and in two independent plantings of each of the three T2 lines. Unlike many early flowering Arabidopsis transgenic lines, which are dwarfed, 35S::G1946 transformants often reached full-size at maturity, and produced large quantities of seeds, although the plants were slightly pale in coloration and had slightly flat leaves compared to wild-type. In addition, 35S::G1946 plants showed an altered response to phosphate deprivation. Seedlings of G1946 overexpressor plants showed more secondary root growth on phosphate-free media, when compared to wild-type control. In a repeat experiment, all three lines showed the phenotype. Overexpression of G1946 in Arabidopsis also resulted in an increase in seed glucosinolate M39501 in T2 lines 1 and 3. An increase in seed oil and a decrease in seed protein was also observed in these two lines. G1946 was ubiquitously expressed, and does not appear to be significantly induced or repressed by any of the biotic and abiotic stress conditions tested at this time, with the exception of cold, which repressed G1946 expression. G1946 can be used to modify flowering time, as well as to improve the plant's performance in conditions of limited phosphate, and to alter seed oil, protein, and glucosinolate composition.

Closely Related Genes from Other Species

A comparison of the amino acid sequence of G1946 with sequences available from GenBank showed strong similarity with plant HSFs of several species (*Lycopersicon peruvianum*, *Medicago truncatula*, *Lycopersicon esculentum*, *Glycine max*, *Solanum tuberosum*, *Oryza sativa* and *Hordeum vulgare* subsp. *vulgare*).

G375: The sequence of G375 (SEQ ID NO:239) was experimentally determined and G375 was analyzed using transgenic plants in which G375 was expressed under the control of the 35S promoter. Overexpression of G375 produced marked effects on leaf development. At early stages of growth, 35S::G375 seedlings developed narrow, upward pointing leaves with long petioles (possibly indicating a disruption in circadian-clock controlled processes or nyctinastic movements). Additionally, some seedlings were noted to have elongated hypocotyls, and some were rather small compared to wild-type controls. Comparable phenotypes were obtained by overexpression of an AP2 family gene, G2113 (SEQ ID NO: 85). Following the switch to flowering, 35S::G375 plants showed reduced fertility, which possibly arose from a failure of stamens to fully elongate. One of the three T2 lines, (#41) was later flowering than wild-type controls, and also developed large numbers of small secondary rosette leaves in the axils of the primary rosette. Although these effects were not noted in the other two lines, the phenotypes obtained in line 41 were somewhat similar to those produced by overexpression of another Z-dof gene, G736 (SEQ ID NO: 211). G375 was expressed in all tissues, although at different levels. It was expressed at low levels in the root and germinating seed, and expressed at high levels in the embryo. The effects of G375 on leaf architecture are of potential interest to the ornamental horticulture industry.

Closely Related Genes from Other Species

G375 showed some homology to non-Arabidopsis plant proteins within the conserved Dof domain.

G1255: The sequence of G1255 (SEQ ID NO: 273) was experimentally determined and G1255 was analyzed using transgenic plants in which G1255 was expressed under the control of the 35S promoter. Plants overexpressing G1255 had

alterations in leaf architecture, a reduction in apical dominance, an increase in seed size, and showed more disease symptoms following inoculation with a low dose of the fungal pathogen *Botrytis cinerea*. G1255 was constitutively expressed and not significantly induced by any conditions tested. On the basis of the phenotypes produced by overexpression of G1255, G1255 can be used to manipulate the plant's defense response to produce pathogen resistance, alter plant architecture, or alter seed size.

Closely Related Genes from Other Species

G1255 showed strong homology to a putative rice zing finger protein represented by sequence AC087181_3. Sequence identity between these two protein extended beyond the conserved domain, and therefore, these genes can be orthologs.

G865: The complete cDNA sequence of G865 (SEQ ID NO: 557) was determined. G865 was ubiquitously expressed in *Arabidopsis* tissues. G865 was analyzed using transgenic plants in which G865 was expressed under the control of the 35S promoter. Plants overexpressing G865 were early flowering, with numerous secondary inflorescence meristems giving them a bushy appearance. G865 overexpressors were more susceptible to infection with a moderate dose of the fungal pathogens *Erysiphe orontii* and *Botrytis cinerea*. In addition, seeds from G865 overexpressing plants showed a trend of increased protein and reduced oil content, although the observed changes were not beyond the criteria used for judging significance except in one line. G865 can be used to control flowering time. G865 can be used to manipulate the defense response in order to generate pathogen-resistant plants. G865 can be used to alter seed oil and protein content of a plant.

Closely Related Genes from Other Species

G865 and other non-*Arabidopsis* AP2/EREBP proteins were similar within the conserved AP2 domain.

G2509: G2509 (SEQ ID NO: 23) was studied using transgenic plants in which the gene was expressed under the control of the 35S promoter. Overexpression of G2509 caused multiple alterations in plant growth and development, most notably, altered branching patterns, and a reduction in apical dominance, giving the plants a

shorter, more bushy stature than wild type. Twenty 35S::G2509 primary transformants were examined; at early stages of rosette development, these plants displayed a wild-type phenotype. However, at the switch to flowering, almost all T1 lines showed a marked loss of apical dominance and large numbers of secondary shoots developed from axils of primary rosette leaves. In the most extreme cases, the shoots had very short internodes, giving the inflorescence a very bushy appearance. Such shoots were often very thin and flowers were relatively small and poorly fertile. At later stages, many plants appeared very small and had a low seed yield compared to wild type. In addition to the effects on branching, a substantial number of 35S::G2509 primary transformants also flowered early and had buds visible several days prior to wild type. Similar effects on inflorescence development were noted in each of three T2 populations examined. The branching and plant architecture phenotypes observed in 35S::G2509 lines resemble phenotypes observed for three other AP2/EREBP genes: G865 (SEQ ID NO: 557), G1411 (SEQ ID NO: 3), and G1794 (SEQ ID NO: 13). G2509, G865, and G1411 form a small clade within the large AP2/EREBP family, and G1794, although not belonging to the clade, is one of the AP2/EREBP genes closest to it in the phylogenetic tree. It is thus likely that all these genes share a related function, such as affecting hormone balance. Overexpression of G2509 in Arabidopsis resulted in an increase in alpha-tocopherol in seeds in T2 lines 5 and 11. G2509 was ubiquitously expressed in Arabidopsis plant tissue. G2509 expression levels were altered by a variety of environmental or physiological conditions. G2509 can be used to manipulate plant architecture and development. G2509 can be used to alter tocopherol composition. Tocopherols have anti-oxidant and vitamin E activity. G2509 can be useful in altering flowering time. A wide variety of applications exist for systems that either lengthen or shorten the time to flowering.

Closely Related Genes from Other Species

G2509 showed some sequence similarity with known genes from other plant species within the conserved AP2/EREBP domain.

G2347: G2347 (SEQ ID NO: 1119) was analyzed using transgenic plants in which G2347 was expressed under the control of the 35S promoter. Overexpression of G2347 markedly reduced the time to flowering in Arabidopsis. This phenotype

was apparent in the majority of primary transformants and in all plants from two out of the three T2 lines examined. Under continuous light conditions, 35S::G2347 plants formed flower buds up a week earlier than wild type. Many of the plants were rather small and spindly compared to controls. To demonstrate that overexpression of G2347 could induce flowering under less inductive photoperiods, two T2 lines were re-grown in 12 hour conditions; again, all plants from both lines bolted early, with some initiating flower buds up to two weeks sooner than wild-type. As determined by RT-PCR, G2347 was highly expressed in rosette leaves and flowers, and to much lower levels in embryos and siliques. No expression of G2347 was detected in the other tissues tested. G2347 expression was repressed by cold, and by auxin treatments and by infection by Erysiphe. G2347 is also highly similar to the Arabidopsis protein G2010 (SEQ ID NO: 1121). The level of homology between these two proteins suggested they could have similar, overlapping, or redundant functions in Arabidopsis. In support of this hypothesis, overexpression of both G2010 and G2347 resulted in early flowering phenotypes in transgenic plants.

Closely Related Genes from Other Species

The closest relative to G2347 is the Antirrhinum protein, SBP2 (CAA63061). The similarity between these two proteins is extensive enough to suggest they might have similar functions in a plant.

G988: G988 (SEQ ID NO: 43) was analyzed using transgenic plants in which G988 was expressed under the control of the 35S promoter. Plants overexpressing G988 had multiple morphological phenotypes. The transgenic plants were generally smaller than wild-type plants, had altered leaf, inflorescence and flower development, altered plant architecture, and altered vasculature. In one transgenic line overexpressing G988 (line 23), an increase in the seed glucosinolate M39489 was observed. The phenotype of plants overexpressing G988 was wild-type in all other assays performed. In wild-type plants, G988 was expressed primarily in flower and silique tissue, but was also present at detectable levels in all other tissues tested. Expression of G988 was induced in response to heat treatment, and repressed in response to infection with Erysiphe. Based on the observed morphological phenotypes of the transgenic plants, G988 can be used to create plants with larger flowers. This can have value in the ornamental horticulture industry. The reduction

in the formation of lateral branches suggests that G988 can have utility on the forestry industry. The Arabidopsis plants overexpressing G988 also had reduced fertility. This can be a desirable trait in some instances, as it can be exploited to prevent or minimize the escape of GMO (genetically modified organism) pollen into the environment.

Closely Related Genes from Other Species

The amino acid sequence for the *Capsella rubella* hypothetical protein represented by GenBank accession number CRU303349 was significantly identical to G988 outside of the SCR conserved domains. The *Capsella rubella* hypothetical protein is 90% identical to G988 over a stretch of roughly 450 amino acids. Therefore, it is likely that the *Capsella rubella* gene is an ortholog of G988.

G2346: G2346 (SEQ ID NO: 459) was analyzed using transgenic plants in which the gene was expressed under the control of the 35S promoter. 35S::G2346 seedlings from all three T2 populations had slightly larger cotyledons and appeared somewhat more advanced than controls. This indicated that the seedlings developed more rapidly than the control plants. At later stages, however, G2346 overexpressing plants showed no consistent differences from control plants. The phenotype of these transgenic plants was wild-type in all other assays performed. According to RT-PCR analysis, G2346 is expressed ubiquitously.

Closely Related Genes from Other Species

G2346 shows some sequence similarity with known genes from other plant species within the conserved SBP domain.

G1354: The complete sequence of G1354 (SEQ ID NO: 285) was determined. G1354 was analyzed using transgenic plants in which G1354 was expressed under the control of the 35S promoter. Overexpression of G1354 produced highly deleterious effects on growth and development. Only three 35S::G1354 T1 plants were obtained; all were extremely tiny and slow developing. After three weeks of growth, each of the plants comprised a completely disorganized mass of leaves and root that had no clear axis of growth. Since these individuals would not have survived transplantation to soil, they were harvested for RT-PCR analysis; all three plants showed moderate

levels of G1354 overexpression compared to whole wild-type seedlings of an equivalent size. Only a very small number of transformants were obtained from two selection attempts on separate batches of T0 seed. Usually between 15 and 120 transformants are obtained from each aliquot of 300 mg T0 seed from wild-type plants. The low transformation frequency obtained in this experiment suggests that high levels of G1354 overexpression might have completely lethal effects and prevent transformed seeds from germinating. As determined by RT-PCR, G1354 was uniformly expressed in all tissues and under all conditions tested in RT-PCR. However, the gene was repressed in leaf tissue in response to Erysiphe infection.

Closely Related Genes from Other Species

G1354 is closely related to a NAM protein encoded by polynucleotide from rice (AC005310). Similarity between G1354 and this rice protein extends beyond the signature motif of the family to a level that would suggest the genes are orthologs.

G1063: G1063 (SEQ ID NO: 119) is a member of a clade of highly related HLH/MYC proteins that also includes G779 (SEQ ID NO: 113), G1499 (SEQ ID NO: 7), G2143 (SEQ ID NO: 129), and G2557 (SEQ ID NO: 133). All of these genes caused similar pleiotropic phenotypic effects when overexpressed, the most striking of which was the production of ectopic carpelloid tissue. These genes can be considered key regulators of carpel development. A spectrum of developmental alterations was observed amongst 35S::G1063 primary transformants and the majority were markedly small, dark green, and had narrow curled leaves. The most severely affected individuals were completely sterile and formed highly abnormal inflorescences; shoots often terminated in pin-like structures, and flowers were replaced by filamentous carpelloid structures. In other cases, flowers showed internode elongation between floral whorls, with a central carpel protruding on a pedicel-like organ. Additionally, lateral branches sometimes failed to develop and tiny patches of carpelloid tissue formed at axillary nodes of the inflorescence. In lines with an intermediate phenotype, flowers contained defined whorls of organs, but sepals were converted to carpelloid structures or displayed patches of carpelloid tissue. In contrast, lines with a weak phenotype developed relatively normal flowers and produced a reasonable quantity of seed. Such plants were still distinctly smaller than wild-type controls. Since the strongest 35S::G1063 lines were sterile, three lines

with a relatively weak phenotype, that had produced sufficient seed for biochemical and physiological analysis, were selected for further study. Two of the T2 populations (T2-28,37) were clearly small, darker green and possessed narrow leaves compared to wild type. Plants from one of these populations (T2-28) also produced occasional branches with abnormal flowers like those seen in the T1. The final T2 population (T2-30) displayed a very mild phenotype. Overexpression of G1063 in *Arabidopsis* resulted in a decrease in seed oil content in T2 lines 28 and 37. No altered phenotypes were detected in any of the physiological assays, except that the plants were noted to be somewhat small and produce anthocyanin when grown in Petri plates. G1063 was expressed at low to moderate levels in roots, flowers, rosette leaves, embryos, and germinating seeds, but was not detected in shoots or siliques. It was induced by auxin. G1063 can be used to manipulate flower form and structure or plant fertility. One application for manipulation of flower structure can be in the production of saffron, which is derived from the stigmas of *Crocus sativus*. G1063 has utility in manipulating seed oil and protein content.

Closely Related Genes from Other Species

G1063 protein shared extensive homology in the basic helix loop helix region with a protein sequence encoded by Glycine max cDNA clone (AW832545) as well as a tomato root, plants pre-anthesis *Lycopersicon esculentum* cDNA (BE451174).

G2143: G2143 (SEQ ID NO: 129) is a member of a clade of highly related HLH/MYC proteins that also includes G779 (SEQ ID NO: 113), G1063 (SEQ ID NO: 119), G1499 (SEQ ID NO: 7), and G2557 (SEQ ID NO: 133). All of these genes caused similar pleiotropic phenotypic effects when overexpressed, the most striking of which was the production of ectopic carpelloid tissue. These genes can be considered key regulators of carpel development. Twelve out of twenty 35S::G2143 T1 lines showed a very severe phenotype; these plants were markedly small and had narrow, curled, dark-green leaves. Such individuals were completely sterile and formed highly abnormal inflorescences; shoots often terminated in pin-like structures, and flowers were replaced by filamentous carpelloid structures, or a fused mass of carpelloid tissue. Furthermore, lateral branches usually failed to develop, and tiny patches of stigmatic tissue often formed at axillary nodes of the inflorescence. Strongly affected plants displayed the highest levels of transgene expression

(determined by RT-PCR). The remaining T1 lines showed lower levels of G2143 overexpression; these plants were still distinctly smaller than wild type, but had relatively normal inflorescences and produced seed. Since the strongest 35S::G2143 lines were sterile, three lines with a relatively weak phenotype, that had produced sufficient seed for biochemical analysis, were selected for further study. T2-11 plants displayed a very mild phenotype and had somewhat small, narrow, dark green leaves. The other two T2 populations, however, appeared wild-type, suggesting that transgene activity might have been reduced between the generations. Reduced seedling vigor was noted in the physiological assays. G2143 expression was detected at low levels in flowers and siliques, and at higher levels in germinating seed. G2143 can be used to manipulate flower form and structure or plant fertility. One application for manipulation of flower structure can be in the production of saffron, which is derived from the stigmas of *Crocus sativus*.

Closely Related Genes from Other Species

G2143 protein shared extensive homology in the basic helix loop helix region with a protein encoded by Glycine max cDNA clones (AW832545, BG726819 and BG154493) and a *Lycopersicon esculentum* cDNA clone (BE451174). There was lower homology outside of the region.

G2557: G2557 (SEQ ID NO: 133) is a member of a clade of highly related HLH/MYC proteins that also includes G779 (SEQ ID NO: 113), G1063 (SEQ ID NO: 119), G1499 (SEQ ID NO: 7), and G2143 (SEQ ID NO: 129). All of these genes caused similar pleiotropic phenotypic effects when overexpressed, the most striking of which was the production of ectopic carpelloid tissue. These genes can be considered key regulators of carpel development. The flowers of 35S::G2557 primary transformants displayed patches of stigmatic papillae on the sepals, and often had rather narrow petals and poorly developed stamens. Additionally, carpels were also occasionally held outside of the flower at the end of an elongated pedicel like structure. As a result of such defects, 35S::G2557 plants often showed very poor fertility and formed small wrinkled siliques. In addition to such floral abnormalities, the majority of primary transformants were also small and darker green in coloration than wild type. Approximately one third of the T1 plants were extremely tiny and completely sterile. Three T1 lines (#7,9,12), that had produced some seeds, and

showed a relatively weak phenotype, were chosen for further study. All three of the T2 populations from these lines contained plants that were distinctly small, had abnormal flowers, and were poorly fertile compared to controls. Stigmatic tissue was not noted on the sepals of plants from these three T2 lines. Another line (#4) that had shown a moderately strong phenotype in the T1 was sown for only morphological analysis in the T2 generation. T2-4 plants were small, dark green, and produced abnormal flowers with ectopic stigmatic tissue on the sepals, as had been seen in the parental plant. G2557 expression was detected at low to moderate levels in all tissues tested except shoots. It was induced by cold, heat, and salt, and repressed by pathogen infection. G1063 can be used to manipulate flower form and structure or plant fertility. One application for manipulation of flower structure can be in the production of saffron, which is derived from the stigmas of *Crocus sativus*.

Closely Related Genes from Other Species

G2557 protein shows extensive sequence similarity in the region of basic helix loop helix with a protein encoded by Glycine max cDNA clone (BE347811).

G2430: The complete sequence of G2430 (SEQ ID NO: 697) was determined. G2430 is a member of the response regulator class of GARP proteins (ARR genes), although one of the two conserved aspartate residues characteristic of response regulators is not present. The second aspartate, the putative phosphorylated site, is retained so G2430 can have response regulator function. G2430 is specifically expressed in embryo and silique tissue. In morphological analyses, plants overexpressing G2430 showed more rapid growth than control plants at early stages, and in two of three lines examined produced large, flat leaves. Early flowering was observed for some lines, but this effect was inconsistent between plantings. G2430 can regulate plant growth. Overexpression of G2430 in *Arabidopsis* also resulted in seedlings that are slightly more tolerant to heat in a germination assay. Seedlings from G2430 overexpressing transgenic plants were slightly greener than the control seedlings under high temperature conditions. In a repeat experiment on individual lines, G2430 line 15 showed the strongest heat tolerant phenotype. G2430 can be useful to promote faster development and reproduction in plants.

Closely Related Genes from Other Species

G2430 had some similarity within of the conserved GARP and response-regulator domains to non-Arabidopsis proteins.

G1478: The sequence of G1478 (SEQ ID NO: 831) was determined and G1478 was analyzed using transgenic plants in which G1478 was expressed under the control of the 35S promoter. Plants overexpressing G1478 had a general delay in progression through the life cycle, in particular a delay in flowering time. G1478 is expressed at higher levels in flowers, rosettes and embryos but otherwise expression is constitutive. Based on the phenotypes produced through G1478 overexpression, G1478 can be used to manipulate the rate at which plants grow, and flowering time.

Closely Related Genes from Other Species

G1478 shows some homology to non-Arabidopsis proteins within the conserved domain.

G681: G681 (SEQ ID NO: 579) was analyzed using transgenic plants in which the gene was expressed under the control of the 35S promoter. Approximately half of the 35S::G681 primary transformants were markedly small and formed narrow leaves compared to controls. These plants often produced thin inflorescence stems, had rather poorly formed flowers with low pollen production, and set few seeds. Three T1 lines with relatively weak phenotypes, which had produced reasonable quantities of seed, were selected for further study. Plants from one of the T2 populations were noted to be slightly small, but otherwise the T2 lines displayed no consistent differences in morphology from controls. In leaves of two of the T2 lines, overexpression of G681 resulted in an increase in the percentage of the glucosinolate M39480. According to RT-PCR analysis, G681 expression was detected at very low levels in flower and rosette leaf tissues. G681 was induced by drought stress. G681 can be used to alter glucosinolate composition in plants. Increases or decreases in specific glucosinolates or total glucosinolate content are desirable depending upon the particular application. For example: (1) Glucosinolates are undesirable components of the oilseeds used in animal feed, since they produce toxic effects. Low-glucosinolate varieties of canola have been developed to combat this problem. (2) Some glucosinolates have anti-cancer activity; thus, increasing the levels or

composition of these compounds might be of interest from a nutraceutical standpoint. (3) Glucosinolates form part of a plants natural defense against insects. Modification of glucosinolate composition or quantity could therefore afford increased protection from predators. Furthermore, in edible crops, tissue specific promoters can be used to ensure that these compounds accumulate specifically in tissues, such as the epidermis, which are not taken for consumption.

Closely Related Genes from Other Species

G681 shows some sequence similarity with known genes from other plant species within the conserved Myb domain.

G878: G878 (SEQ ID NO: 611) was studied using transgenic plants in which the gene was expressed under the control of the 35S promoter. Analysis of primary transformants revealed that overexpression of G878 delays the onset of flowering in *Arabidopsis*. 11/20 of the 35S::G878 T1 plants flowered approximately one week later than wild type under continuous light conditions. These plants were also darker green, had shorter stems, and senesced later than controls. G878 was ubiquitously expressed. G878 can be used to modify flowering time and senescence, and a wide variety of applications exist for systems that either lengthen or shorten the time to flowering.

Closely Related Genes from Other Species

G878 was highly related to other WRKY proteins from a variety of plant species, such as the *Nicotiana tabacum* DNA-binding protein 2 (WRKY2) (AF096299), and a *Cucumis sativus* SPF1-like DNA-binding protein (L44134).

G374: G374 (SEQ ID NO: 47) was expressed at low levels throughout the plant and was induced by salicylic acid. G374 was investigated using lines carrying a T-DNA insertion in this gene. The T-DNA insertion was approximately three quarters of the way into the protein coding sequence and should result in a null mutation. Homozygosity for a T-DNA insertion within G374 caused lethality at early stages of embryo development. In an initial screen for G374 knockouts, heterozygous plants were identified. Seed from those individuals was sown to soil and eleven plants were PCR-screened to identify homozygotes. No homozygotes were obtained;

6 of the progeny were heterozygous whilst the other 5 were wild type. This raised the prospect that homozygosity for the G374 insertion was lethal. To examine this possibility further, heterozygous KO.G374 plants were re-grown. These individuals looked wild type, but their siliques were examined for seed abnormalities. When green siliques were dissected, around 25% of developing seeds were white or aborted. Embryos from these siliques were cleared using Hoyer's solution, and examined under the microscope. It was apparent that embryos from the white seeds had arrested at early (globular or heart) stages of development, whilst embryos from the normal seeds were fully developed. Such arrested or aborted seeds most likely represented homozygotes for the G374 insertion. To support this conclusion, seed was collected from heterozygous plants and sown to kanamycin plates (the T-DNA insertion carried the NPT marker gene). Of the seedlings that germinated, 160 were kanamycin resistant and 107 were kanamycin sensitive. These data more closely fitted a 2:1 (χ^2 , 1df, = 5.5, $0.05 > P > 0.01$) than a 3:1 (χ^2 , 1df, = 32, $P < 0.001$) ratio. Such a segregation ratio suggested that a homozygous class of kanamycin resistant seedlings was absent from the progeny of KO.G374 plant. G374 can be a herbicide target.

Closely Related Genes from Other Species

Similar sequences to G374 are present in tomato and *Medicago truncatula*, and these sequences can be orthologs.

Example VIII: Identification of Homologous Sequences

Homologous sequences from *Arabidopsis* and plant species other than *Arabidopsis* were identified using database sequence search tools, such as the Basic Local Alignment Search Tool (BLAST) (Altschul et al. (1990) *J. Mol. Biol.* 215:403-410; and Altschul et al. (1997) *Nucl. Acid Res.* 25: 3389-3402). The tblastx sequence analysis programs were employed using the BLOSUM-62 scoring matrix (Henikoff, S. and Henikoff, J. G. (1992) *Proc. Natl. Acad. Sci. USA* 89: 10915-10919).

Identified non-*Arabidopsis* sequences homologous to the *Arabidopsis* sequences are provided in Table 5. The percent sequence identity among these sequences can be as low as 47%, or even lower sequence identity. The entire NCBI GenBank database was filtered for sequences from all plants except *Arabidopsis thaliana* by selecting all entries in the NCBI GenBank database associated with NCBI

taxonomic ID 33090 (Viridiplantae; all plants) and excluding entries associated with taxonomic ID 3701 (*Arabidopsis thaliana*). These sequences are compared to sequences representing genes of SEQ IDs NOs:2 - 2N, where N = 2-561, using the Washington University TBLASTX algorithm (version 2.0a19MP) at the default settings using gapped alignments with the filter "off". For each gene of SEQ IDs NOs:2 - 2N, where N = 2-561, individual comparisons were ordered by probability score (P-value), where the score reflects the probability that a particular alignment occurred by chance. For example, a score of 3.6×10^{-40} is 3.6×10^{-40} . In addition to P-values, comparisons were also scored by percentage identity. Percentage identity reflects the degree to which two segments of DNA or protein are identical over a particular length. Examples of sequences so identified are presented in Table 5. Homologous or orthologous sequences are readily identified and available in GenBank by Accession number (Table 5; Test sequence ID). The identified homologous polynucleotide and polypeptide sequences and homologues of the *Arabidopsis* polynucleotides and polypeptides may be orthologs of the *Arabidopsis* polynucleotides and polypeptides (TBD: to be determined).

Example IX Introduction of polynucleotides into dicotyledonous plants

SEQ ID NOs:1-(2N - 1), wherein N = 2-561, paralogous, orthologous, and homologous sequences recombined into pMEN20 or pMEN65 expression vectors are transformed into a plant for the purpose of modifying plant traits. The cloning vector may be introduced into a variety of cereal plants by means well-known in the art such as, for example, direct DNA transfer or *Agrobacterium tumefaciens*-mediated transformation. It is now routine to produce transgenic plants using most dicot plants (see Weissbach and Weissbach, (1989) *supra*; Gelvin et al., (1990) *supra*; Herrera-Estrella et al. (1983) *supra*; Bevan (1984) *supra*; and Klee (1985) *supra*). Methods for analysis of traits are routine in the art and examples are disclosed above.

Example X Transformation of Cereal Plants with an Expression Vector

Cereal plants such as corn, wheat, rice, sorghum or barley, may also be transformed with the present polynucleotide sequences in pMEN20 or pMEN65 expression vectors for the purpose of modifying plant traits. For example, pMEN020 may be modified to replace the NptII coding region with the BAR gene of *Streptomyces hygroscopicus* that confers resistance to phosphinothricin. The KpnI

and BglII sites of the Bar gene are removed by site-directed mutagenesis with silent codon changes.

The cloning vector may be introduced into a variety of cereal plants by means well-known in the art such as, for example, direct DNA transfer or *Agrobacterium tumefaciens*-mediated transformation. It is now routine to produce transgenic plants of most cereal crops (Vasil, I., Plant Molec. Biol. 25: 925-937 (1994)) such as corn, wheat, rice, sorghum (Cassas, A. et al., Proc. Natl. Acad Sci USA 90: 11212-11216 (1993) and barley (Wan, Y. and Lemeaux, P. Plant Physiol. 104:37-48 (1994). DNA transfer methods such as the microprojectile can be used for corn (Fromm. et al. Bio/Technology 8: 833-839 (1990); Gordon-Kamm et al. Plant Cell 2: 603-618 (1990); Ishida, Y., Nature Biotechnology 14:745-750 (1990)), wheat (Vasil, et al. Bio/Technology 10:667-674 (1992) ; Vasil et al., Bio/Technology 11:1553-1558 (1993); Weeks et al., Plant Physiol. 102:1077-1084 (1993)), rice (Christou Bio/Technology 9:957-962 (1991); Hiei et al. Plant J. 6:271-282 (1994); Aldemita and Hodges, Planta 199:612-617; Hiei et al., Plant Mol Biol. 35:205-18 (1997)). For most cereal plants, embryogenic cells derived from immature scutellum tissues are the preferred cellular targets for transformation (Hiei et al., Plant Mol Biol. 35:205-18 (1997); Vasil, Plant Molec. Biol. 25: 925-937 (1994)).

Vectors according to the present invention may be transformed into corn embryogenic cells derived from immature scutellar tissue by using microprojectile bombardment, with the A188XB73 genotype as the preferred genotype (Fromm, et al., Bio/Technology 8: 833-839 (1990); Gordon-Kamm et al., Plant Cell 2: 603-618 (1990)). After microprojectile bombardment the tissues are selected on phosphinothricin to identify the transgenic embryogenic cells (Gordon-Kamm et al., Plant Cell 2: 603-618 (1990)). Transgenic plants are regenerated by standard corn regeneration techniques (Fromm, et al., Bio/Technology 8: 833-839 (1990); Gordon-Kamm et al., Plant Cell 2: 603-618 (1990)).

The plasmids prepared as described above can also be used to produce transgenic wheat and rice plants (Christou, Bio/Technology 9:957-962 (1991); Hiei et al., Plant J. 6:271-282 (1994); Aldemita and Hodges, Planta 199:612-617 (1996); Hiei et al., Plant Mol Biol. 35:205-18 (1997)) that coordinately express genes of

interest by following standard transformation protocols known to those skilled in the art for rice and wheat Vasil, et al. Bio/Technology 10:667-674 (1992) ; Vasil et al., Bio/Technology 11:1553-1558 (1993); Weeks et al., Plant Physiol. 102:1077-1084 (1993)), where the bar gene is used as the selectable marker.

All references, publications, patent documents, web pages, and other documents cited or mentioned herein are hereby incorporated by reference in their entirety for all purposes. Although the invention has been described with reference to specific embodiments and examples, it should be understood that one of ordinary skill can make various modifications without departing from the spirit of the invention. The scope of the invention is not limited to the specific embodiments and examples provided.

We claim:

1. A transgenic plant comprising a recombinant polynucleotide having a nucleotide sequence selected from the group consisting of:
 - (a) a nucleotide sequence encoding a polypeptide comprising a sequence selected from those of SEQ ID NOs: 860, 802, 240, 274, 558, 24, 1120, 44, 460, 286, 120, 130, 134, 698, 832, 580, 612, and 48, or a complementary nucleotide sequence thereof;
 - (b) a nucleotide sequence of SEQ ID NOs: 859, 801, 239, 273, 557, 23, 1119, 43, 459, 285, 119, 129, 133, 697, 831, 579, 611, 47, or a complementary nucleotide sequence thereof; and
 - (c) a nucleotide sequence that hybridizes under stringent conditions to a nucleotide sequence of one or more polynucleotides of: (a) or (b).
2. The transgenic plant of claim 1 wherein the transgenic plant possesses an altered trait as compared to another plant, or the transgenic plant exhibits an altered phenotype as compared to another plant, or the transgenic plant expresses an altered level of one or more genes associated with a plant trait as compared to another plant, wherein the other plant does not comprise the recombinant polynucleotide.
3. The transgenic plant of claim 1 wherein the plant possesses an altered trait as compared to another plant wherein the trait is an alteration in one or more physical characteristics selected from the group consisting of: the number of trichomes, fruit and seed size and number, yield of stems, leaves, inflorescences, or roots, stability of seeds during storage, susceptibility of the seed to shattering, root hair length and quantity, internode distances, or the quality of seed coat.
4. The transgenic plant of claim 1 wherein the plant possesses an altered trait as compared to another plant wherein the trait is an alteration in a plant growth characteristic selected from the group consisting of: growth rate, germination rate of seeds, vigor of plants and seedlings, leaf and flower senescence, male sterility, apomixis, flowering time, flower abscission, rate of nitrogen uptake, osmotic sensitivity to soluble sugar concentrations, biomass or transpiration characteristics,

apical dominance, branching patterns, number of organs, organ identity, and organ shape or size.

5. The transgenic plant of claim 1 wherein the plant possesses an altered trait as compared to another plant wherein the trait is an alteration in one or more characteristics selected from the group consisting of protein or oil production, seed protein or oil production, insoluble sugar level, soluble sugar level, and starch composition.
6. The transgenic plant of claim 1, wherein the recombinant polynucleotide comprises a nucleotide sequence encoding a polypeptide of SEQ ID NO:860.
7. The transgenic plant of claim 1, wherein the recombinant polynucleotide comprises a nucleotide sequence encoding a polypeptide of SEQ ID NO:802.
8. The transgenic plant of claim 1, wherein the recombinant polynucleotide comprises a nucleotide sequence encoding a polypeptide of SEQ ID NO:240.
9. The transgenic plant of claim 1, wherein the recombinant polynucleotide comprises a nucleotide sequence encoding a polypeptide of SEQ ID NO:274.
10. The transgenic plant of claim 1, wherein the recombinant polynucleotide comprises a nucleotide sequence encoding a polypeptide of SEQ ID NO:558.
11. The transgenic plant of claim 1, wherein the recombinant polynucleotide comprises a nucleotide sequence encoding a polypeptide of SEQ ID NO:24.
12. The transgenic plant of claim 1, wherein the recombinant polynucleotide comprises a nucleotide sequence encoding a polypeptide of SEQ ID NO:1120.
13. The transgenic plant of claim 1, wherein the recombinant polynucleotide comprises a nucleotide sequence encoding a polypeptide of SEQ ID NO:44.

14. The transgenic plant of claim 1, wherein the recombinant polynucleotide comprises a nucleotide sequence encoding a polypeptide of SEQ ID NO:460.
15. The transgenic plant of claim 1, wherein the recombinant polynucleotide comprises a nucleotide sequence encoding a polypeptide of SEQ ID NO:286.
16. The transgenic plant of claim 1, wherein the recombinant polynucleotide comprises a nucleotide sequence encoding a polypeptide of SEQ ID NO:120.
17. The transgenic plant of claim 1, wherein the recombinant polynucleotide comprises a nucleotide sequence encoding a polypeptide of SEQ ID NO:130.
18. The transgenic plant of claim 1, wherein the recombinant polynucleotide comprises a nucleotide sequence encoding a polypeptide of SEQ ID NO:134.
19. The transgenic plant of claim 1, wherein the recombinant polynucleotide comprises a nucleotide sequence encoding a polypeptide of SEQ ID NO:698.
20. The transgenic plant of claim 1, wherein the recombinant polynucleotide comprises a nucleotide sequence encoding a polypeptide of SEQ ID NO:832.
21. The transgenic plant of claim 1, wherein the recombinant polynucleotide comprises a nucleotide sequence encoding a polypeptide of SEQ ID NO:580.
22. The transgenic plant of claim 1, wherein the recombinant polynucleotide comprises a nucleotide sequence encoding a polypeptide of SEQ ID NO:612.
23. The transgenic plant of claim 1, wherein the recombinant polynucleotide comprises a nucleotide sequence encoding a polypeptide of SEQ ID NO:48.
24. The transgenic plant of claim 1, wherein the recombinant polynucleotide comprises a nucleotide sequence of SEQ ID NO:859.

25. The transgenic plant of claim 1, wherein the recombinant polynucleotide comprises a nucleotide sequence of SEQ ID NO:801.
26. The transgenic plant of claim 1, wherein the recombinant polynucleotide comprises a nucleotide sequence of SEQ ID NO:239.
27. The transgenic plant of claim 1, wherein the recombinant polynucleotide comprises a nucleotide sequence of SEQ ID NO:273.
28. The transgenic plant of claim 1, wherein the recombinant polynucleotide comprises a nucleotide sequence of SEQ ID NO:557.
29. The transgenic plant of claim 1, wherein the recombinant polynucleotide comprises a nucleotide sequence of SEQ ID NO:23.
30. The transgenic plant of claim 1, wherein the recombinant polynucleotide comprises a nucleotide sequence of SEQ ID NO:1119.
31. The transgenic plant of claim 1, wherein the recombinant polynucleotide comprises a nucleotide sequence of SEQ ID NO:43.
32. The transgenic plant of claim 1, wherein the recombinant polynucleotide comprises a nucleotide sequence of SEQ ID NO:459.
33. The transgenic plant of claim 1, wherein the recombinant polynucleotide comprises a nucleotide sequence of SEQ ID NO:285.
34. The transgenic plant of claim 1, wherein the recombinant polynucleotide comprises a nucleotide sequence of SEQ ID NO:119.
35. The transgenic plant of claim 1, wherein the recombinant polynucleotide comprises a nucleotide sequence of SEQ ID NO:129.

36. The transgenic plant of claim 1, wherein the recombinant polynucleotide comprises a nucleotide sequence of SEQ ID NO:133.
37. The transgenic plant of claim 1, wherein the recombinant polynucleotide comprises a nucleotide sequence of SEQ ID NO:697.
38. The transgenic plant of claim 1, wherein the recombinant polynucleotide comprises a nucleotide sequence of SEQ ID NO:831.
39. The transgenic plant of claim 1, wherein the recombinant polynucleotide comprises a nucleotide sequence of SEQ ID NO:579.
40. The transgenic plant of claim 1, wherein the recombinant polynucleotide comprises a nucleotide sequence of SEQ ID NO:611.
41. The transgenic plant of claim 1, wherein the recombinant polynucleotide comprises a nucleotide sequence of SEQ ID NO:47.
42. The transgenic plant of claim 1, further comprising a constitutive, inducible, or tissue-specific promoter operably linked to said nucleotide sequence.
44. The transgenic plant of claim 1, wherein the plant is selected from the group consisting of: soybean, wheat, corn, potato, cotton, rice, oilseed rape, sunflower, alfalfa, sugarcane, turf, banana, blackberry, blueberry, strawberry, raspberry, cantaloupe, carrot, cauliflower, coffee, cucumber, eggplant, grapes, honeydew, lettuce, mango, melon, onion, papaya, peas, peppers, pineapple, pumpkin, spinach, squash, sweet corn, tobacco, tomato, watermelon, mint and other labiates, rosaceous fruits, and vegetable brassicas.
44. The transgenic plant of claim 1 wherein the encoded polypeptide is expressed and regulates transcription of a gene.

45. A method of using the transgenic plant of claim 1 to grow a progeny plant from a parent plant, the method comprising crossing the transgenic plant with another plant, selecting seed, and growing the progeny plant from the seed.

46. An isolated or recombinant polynucleotide comprising a nucleotide sequence selected from the group consisting of:

- (a) a nucleotide sequence encoding a polypeptide comprising a sequence selected from SEQ ID NOs: 240, 274, 558, 286, 698, and 832, or a complementary nucleotide sequence thereof;
- (b) a nucleotide sequence of SEQ ID NOs: 239, 273, 557, 285, 697, 831, or a complementary nucleotide sequence thereof; and
- (c) a nucleotide sequence that hybridizes under stringent conditions to a nucleotide sequence of one or more of: (a) or (b).

47. The isolated or recombinant polynucleotide of claim 46, wherein the recombinant polynucleotide comprises a nucleotide sequence encoding a polypeptide of SEQ ID NO:240.

48. The isolated or recombinant polynucleotide of claim 46, wherein the recombinant polynucleotide comprises a nucleotide sequence encoding a polypeptide of SEQ ID NO:274.

49. The isolated or recombinant polynucleotide of claim 46, wherein the recombinant polynucleotide comprises a nucleotide sequence encoding a polypeptide of SEQ ID NO:558.

50. The isolated or recombinant polynucleotide of claim 46, wherein the recombinant polynucleotide comprises a nucleotide sequence encoding a polypeptide of SEQ ID NO:286.

51. The isolated or recombinant polynucleotide of claim 46, wherein the recombinant polynucleotide comprises a nucleotide sequence encoding a polypeptide of SEQ ID NO:698.

52. The isolated or recombinant polynucleotide of claim 46, wherein the recombinant polynucleotide comprises a nucleotide sequence encoding a polypeptide of SEQ ID NO:832.

53. The isolated or recombinant polynucleotide of claim 46, wherein the recombinant polynucleotide comprises a nucleotide sequence of SEQ ID NO:239.

54. The isolated or recombinant polynucleotide of claim 46, wherein the recombinant polynucleotide comprises a nucleotide sequence of SEQ ID NO:273.

55. The isolated or recombinant polynucleotide of claim 46, wherein the recombinant polynucleotide comprises a nucleotide sequence of SEQ ID NO:557.

56. The isolated or recombinant polynucleotide of claim 46, wherein the recombinant polynucleotide comprises a nucleotide sequence of SEQ ID NO:285.

57. The isolated or recombinant polynucleotide of claim 46, wherein the recombinant polynucleotide comprises a nucleotide sequence of SEQ ID NO:697.

58. The isolated or recombinant polynucleotide of claim 46, wherein the recombinant polynucleotide comprises a nucleotide sequence of SEQ ID NO:831.

59. The isolated or recombinant polynucleotide of claim 46, further comprising a constitutive, inducible, or tissue-specific promoter operably linked to the nucleotide sequence.

60. The isolated or recombinant polynucleotide of claim 46 wherein the encoded polypeptide is expressed and regulates transcription of a gene.

61. A vector comprising the isolated or recombinant polynucleotide of claim 46.

62. A host cell comprising the vector of claim 61.

63. A method of using the isolated or recombinant polynucleotide of claim 46 for producing a plant having a modified trait, the method comprising selecting a polynucleotide that encodes a polypeptide, inserting the polynucleotide into an expression vector, introducing the vector into a plant or a cell of a plant to overexpress the polypeptide, thereby producing a modified plant, and selecting a modified plant for a modified trait.

64. The method of claim 63 wherein the plant possesses a modified trait as compared to another plant wherein the trait is an alteration in one or more physical characteristics selected from the group consisting of: the number of trichomes, fruit and seed size and number, yield of stems, leaves, inflorescences, or roots, stability of seeds during storage, susceptibility of the seed to shattering, root hair length and quantity, internode distances, or the quality of seed coat.

65. The method of claim 63 wherein the plant possesses a modified as compared to another plant wherein the trait is an alteration in a plant growth characteristic selected from the group consisting of: growth rate, germination rate of seeds, vigor of plants and seedlings, leaf and flower senescence, male sterility, apomixis, flowering time, flower abscission, rate of nitrogen uptake, osmotic sensitivity to soluble sugar concentrations, biomass or transpiration characteristics, apical dominance, branching patterns, number of organs, organ identity, and organ shape or size.

66. The method of claim 63 wherein the plant possesses a modified trait as compared to another plant wherein the trait is an alteration in one or more characteristics selected from the group consisting of protein or oil production, seed protein or oil production, insoluble sugar level, soluble sugar level, and starch composition.

67. A modified plant produced by the method of claim 63.

68. A method of using the plant of claim 67 to grow a progeny plant from a parent plant, the method comprising crossing the transgenic plant with another plant, selecting seed, and growing the progeny plant from the seed.

69. The plant produced by the method of claim 68.

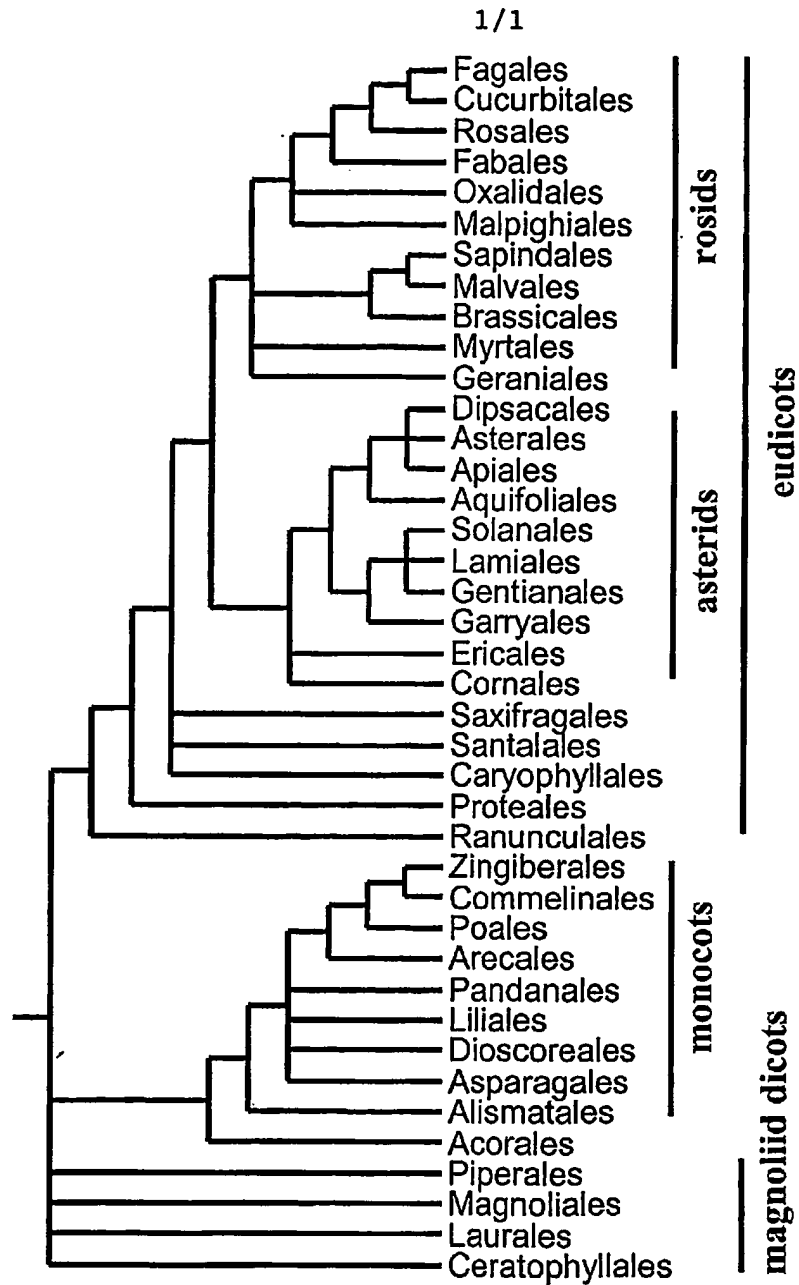


Figure 1

SEQUENCE LISTING

<110> Mendel Biotechnology, Inc.
 Ratcliffe, Oliver
 Riechmann, Jose Luis
 Adam, Luc J.
 Dubell, Arnold T.
 Heard, Jacqueline E.
 Pilgrim, Marsha L.
 Jiang, Cai-Zhong
 Reuber, T. Lynne
 Creelman, Robert A.
 Pineda, Omaira
 Yu, Guo-Liang
 Broun, Pierre E.

<120> YIELD-RELATED POLYNUCLEOTIDES AND
 POLYPEPTIDES IN PLANTS

<130> 514442002041

<150> 60/310,847

<151> 2001-08-09

<150> 60/336,049

<151> 2001-11-19

<150> 60/338,692

<151> 2001-12-11

<150> 10/171,468

<151> 2002-06-14

>G1275 (58..579)

CCAAGAAAAGGGAAGATCACGCATTCCTTATAGGCGTAATTCGTAAATAGTGGTGAGTATG
 AATGATGCAGACACAAACTTGGGGAGTAGTTTTAGCGATGATACTCACTCTGTGTTGAG
 TTTCCGGAGCTAGACTTGTCTAGATGAATGGATGGATGATGATCTTGTGTCTGCGGTTTCC
 GGGATGAATCAGTCTTATGGTTATCAGACTAGTGTGCTGGTGGTCTTTATCTCAGGT
 TCTTCTAGCTGTTTCAGTCATCCTGAATCTCCAAGTACCAAACTTATGTTGCTGCTACA
 GCCACTGCTTCTGCCGACAACCAAAACAAGAAAGAAAAGAAAAAATTAAAGGGAGAGTT
 GCGTTCAAGACACGGTCCGAGGTGGAAGTGCTTGACGACGGGTTCAAGTGGAGAAAGTAT
 GGGAAGAAGATGGTGAAGAACAGCCACATCCAAGAACTACTACAAATGTTTCAAGTTGAT
 GGCTGTCCCGTGAAGAAAAGGTTGAACGAGACAGAGATGATCCGAGCTTTGTGATAACA
 ACTTACGAGGGTTCCCAACAATCACTCAAGCATGAACCTAAGACTCGAAGCTAAGGCTCAAGG
 CGACCATGCTATATTCAGCACATCTTATTTCTATGGTTACGAACGATACTTAAAACTGC
 TTCTAGTTCTTTATATCCATTGTAACTGGTTGCAGGTTCAAAATTTTGAGAGGTTTAT
 GACATTCTAAATCTGTAGTACTTATATA

>G1275 Amino Acid Sequence (domain in AA coordinates: 113-169)
 MNDADTNLGSFSDDTHSVFEPPELDLSDEWMDLVSASVSGMNSYGYQTSVDVAGALFS
 GSSSCFSPHESPSTKTYVAATASADNQNKKKKIKGRVAFKTRSEVEVLDDGFKWRK
 YGKMKVNKNSPHPRNYKCSVDGCPVKRVERDRDDPSFVITTYEGSHNHSSMN*

>G1411 (110..856)

TAAAGAAAACTGAACAACCCTAAAGTACTGTATAAATCCTATATCAAATTTTTTTTTTG
 GAAGAAAAGGCTATATTTAAAGAAAATCAAGCAAAAGTAGATCCTCGGATGTATGGGAA
 GAGGCCTTTTGAGAGGTGATGAATCTGAAGAAAGGGAAGAAGATGAGAAGTTGTTCCCGGT
 CTTCTCGGCCCGATCTCAACACGACATGCGTGTTATGGTCTCGGCCTTGACTCAAGTAAT

CGGAAACCAACAAAGCAAATCTCATGATAACATCAGCTCTATTGATGATAACTATCCTTC
TGTGTATAATCCACAAGACCCTAATCAACAAGTTGCGCCTACTCATCAAGACCAAGGGGA
CTTGAGGAGGAGACATTATAGAGGTGTAAAGCAAAGGCCATGGGGAAGTGGGCAGCTGA
AATCCGAGACCCAAAAAAGGCGGCACGTGTGTGGCTCGGGACATTTGAAACCGCTGAATC
TGCGGCCTTAGCTTATGATGAAGCAGCCCTAAAGTTCAAAGGAAGCAAAGCAAACTCAA
TTTCCCGGAGAGGGTTTCTAGCTTGAAGTAACCTACATATTACTCCTCCAACCAAATTCC
ACAAATGGAACCACAAAGTATACCGAATAATCAATACTATCATGATGCGAGTAGTGG
TGATATGCTAAGTTTAAATTTGGGCGGTGGGTATGGGAGTGGTACCGGATATTCAATGTC
TCATGATAATAGTACTACGACTGCTGTACAACCTTCTTCGTTCTCTGGTGGCTCTTCTAG
GCAACAAGAAGAGCAAGATTATGCCAGATTCTGGCGCTTTGGGGATTCTTCTTCTCTCC
TCAATTCGGGATATTAAATAGGAGATTGATCAGTTACTTGTGATGAAGTAATGATACATT
TCCCGTCAAATTTGAGATGATCATATGCTTCTGAATGTTTTGAGTGTCAATTTTGTCT
TCCGCGTTAAGATTTATTGAACGTGTTTTCTTGTTTTTTGGTTAAAAA

>G1411 Amino Acid Sequence (domain in AA coordinates: 87-154)
MYGKRPFPGDESEEREEDENLFPVFSARSQHDMRVMVSALTQVIGNQQSKSHDNISID
NYPVYNPQDPNQVAPTHQDQGLRRRHRYGVRQRPWGKWAABEIRDPKKAARVWLGTFE
TAESAALAYDEAALKFKGSKAKLNFPERVQLGSNSTYYSSNQIPQMEPQSI PNYNQYYHD
ASSGDMLSFNLGGYGSGTGYSMSHDNSTTTAATSSSSGSSSRQEEQDYARFWRFGDS
SSSPHSYG*

>G1488 (1..996)
ATGGAAGATGAAGCACATGAATCTTCCACACATCTGATTTTGCCGTGATGACCTTTTA
GTTGATTTCTCTAACGATGATGACGAAGAAAACGATGTTGTTGCTGATTCACCACTACC
ACCACCATAACCGACAGCTCTAATCTTCCGCTGCTGATCTTCCAGTTTCCACGGTGAT
GTTCAAGACGGCACTAGCTTCTCCGGTGACCTTTGTATACCTTCTGATGATTTGGCTGAT
GAGTTAGAGTGGCTTTCGAACATTGTGGATGAATCATTTGTCGCTGAAGATGTACACAAG
CTCGAGCTAATATCCGGTTTTAAGAGTGCACCGGACCCGAAATCCGATACCGGAAGCCCG
GAAAACCCGAATAGCAGCAGTCCGATTTTACTACCGACGTTTCTGTACCGGCCAAAGCT
AGAAGCAAACGCTCACGCGCCGCTGCGTGTAATTGGGCCCTCACGTGGGCTTCTCAAGGAA
ACGTTTTACGACAGTCTTTTACCAGGAGAAACCATTTCTCTAGCCAACAACACTTGTCT
CCGCCAACCTCGCCGCTTTGTTGATGGCTCCGCTAGGGAAGCAAGCCGTTGATGGA
GGACACCGACGGAAGAAGGATGTTCTTACCAGGAGTCTGGTGGCGCAGAGGAGAGACGG
TGTCTCCACTGCGCCACGGATAAGACTCCGCAATGGCGACAGGCCCAATGGGCCCAGG
ACGTTGTGCAACGCTTGGCGTGTAGGTACAAATCGGGACGTTTAGTGCCGGAGTATCGG
CCCGCGCGAGTCCGACGTTTGTGCTGGCGAAACACTCAAATCTCATCGGAAAGTTATG
GAGCTCCGGCGACAGAAGGAGATGAGTAGGGCCCATCATGAGTTCATACATCACCATCAC
GGTACGGACACTGCCATGATTTTCGACGTTTTCATCGGACGGTGATGATTACTTGATCCAC
CACAACGTTGGCCAGATTTTACAGACAGCTTATTTGA

>G1488 Amino Acid Sequence (domain in AA coordinates: 221-246)
MEDEAHEFFHTSDFAVDDLVDLFDNDDEENDVVADSTTTTTITDSSNFSAADLPSFHGD
VQDGTFSFGDLCPISDDLDELEWLSNIVDESLSPELVHKLILISGFKSRPDPKSDTGSP
ENPNSSSPIFTDVSVPKARSKRSRAAACNWA SRGLLKETFYDSPFTGETILSSQQHLS
PPTSPPLLMAPLGKKQAVDGGHRRKKDVSSPESGGAEERRCLHCATDKTPQWRTGPMGPK
TLCNACGVRYKSGRLVPEYRPAASPTFVLAKHSNSHRKVMELRRQKEMSRHHEFIHHHH
GTD TAMI F DVSSDGD DYL I H H NVGP DFRQLI*

>G1499 (159..833)
TCGACTCTTAATTGCATCACCAACCTAACCCTCATCATATGATCACCTTTCTTTTTTT
CCTTTAATATATATATATGCTCACACACACATATATATATACATATAAGCATCGCCTC
AAGCATTAATAATTTTACGAACCAACAAACAAAAATTATGAATAATTATAATATGAACC
CATCTCTCTTCCAAAATTACACTTGAACAACATCATCAACAGCAGCAACAACAACA
AGAATGATGATCATCATCAACATAATAATGATCCAATCGGTATGGCCATGGACCAGT
ACACACAGCTCCATATCTTCAATCCTTTCTCTCTTCTCATTTCCTCCTCTCTCTCTT
CCCTCACAACCACCACTTCTCTCCGGAGATCAAGAAGACGACGAAGACGAAGAAGAAC
CTCTAGAGGAACCTCGGTGCTATGAAGGAAATGATGTACAAGATCGCAGCCATGCAATCGG
TTGACATCGACCCAGCAACCGTCAAGAAACCCAAACGCCGTAACGTGAGGATCTCCGACG
ACCTCAGAGTGTGGCGGTAGACATCGCCGTGAGAGAATCAGTGAGAGGATCAGAATTC
TTCAGAGACTCGTGCCAGGTGGCACTAAATGGATACGGCTTCAATGCTCGATGAAGCTA

TACGCTATGTCAAGTTCTTTGAAACGGCAGATCCGGCTACTCAATAATAATACCGGATATA
CTCCTCCGCCGCCGCAAGATCAAGCTTCTCAGGCGGTGACGACGTCATGGGTTTCACCGC
CACCACGCCCAAGTTTCGGCCGTGGGGGAAGAGGAGTAGGAGAATTAATCTAGACAAGAT
GACATTTCCATTAGTAGTAACATAAATTATGCTATAATGTGTGAGTAATGGTGCAATTATG
GA

>G1499 Amino Acid Sequence (domain in AA coordinates: 118-181)
MNNYNMNP SLFQNYTWNNI INSSNNNNKNDH HHQHNDPIGMAMDQYTLHIFNPFSSS
HFPPLSSSLTTTLLSGDQEDDEDEEEPLEELGAMKEMMYKIAAMQSVDIDPATVKKPKR
RNVRI SDDPQSVAARHRRERISERIRILQRLVPGGTKMDTASMLDEAIRYVKFLKRQIRL
LNNNTGYTPPPPQDQASQAVTTSWVSPPPPPSFRGGRGVGELI*

>G1543 (1..828)
ATGATAAACTACTATTTTACGTACATATGCACATACACATATAAACTATATGCTCTATAT
CATATGGATTACGCATGCGTGTGTATGTATAAATATAAAGGCATCGTCACGCTTCAAGTT
TGTCTCTTTTATATTAACTGAGAGTTTTCTCTCAAACCTTTACCTTTTCTTCTTCGATC
CTAGCTCTTAAGAACCCTAATAATTCATTGATCAAATAATGGCGATTTTGCCGGAAAAC
TCTTCAAACCTTGGATCTTACTATCTCCGTTCCAGGCTTCTCTTCATCCCCCTCTCTCCGAT
GAAGGAAGTGGCGGAGGAAGAGACCAGCTAAGGCTAGACATGAATCGGTTACCGTCTGTCT
GAAGACGGAGACGATGAAGAATTCAGTCACGATGATGGCTCTGTCTCTCCCGGAAAGAAA
CTCCGCTTAACAGAGAACAGTCACGCTCTTCTGAAGATAGTTTCAGACAGAATCATACC
CTTAATCCCAAAACAAAGGAAGTACTTGCCAAGCATTTGATGCTACGGCCAAGACAAATT
GAAGTTTGGTTTCAAACCGTAGAGCAAGGAGCAAATTGAAGCAAACCGAGATGGAATGC
GAGTATCTCAAAGGTGGTTTGGTTTCATTAACGGAAGAAAACACAGGCTCCATAGAGAA
GTAGAAGAGCTTAGAGCCATAAAGGTTGGCCCAACAACGGTGAACCTCTGCCTCGAGCCTT
ACTATGTGTCTCTCGCTGCGAGCGAGTTACCCCTGCCGCGAGCCCTTCGAGGGCGGTGGTG
CCGGTTCCGGCTAAGAAAACGTTTCCGCCGCAAGAGCGTGATCGTTGA

>G1543 Amino Acid Sequence (domain in AA coordinates: 135-195)
MIKLLFTYICTYTYKLYALYHMDYACVCMYKYKIVTLQVCLFYIKLRVFLSNFTFSSSI
LALKNPNNSLIKIMAILPENSSNLDLTI SVPGFSSSPLSDEGSGGGRDQLRLDMNRLPSS
EDGDDEEFSHDDGSAPPRKRLRLTREQSRLL EDSFRQNH TLNPKQKEVLAKHLMRLPRQI
EVWFQNRARSKLKQTEMECEYLKRWFGSLTEENHRLHREVEELRAIKVGPTTVNSASSL
TMCPRCERVTPAASPSRAVVPVPAKKTFFPQERDR*

>G1635 (1..1164)
ATGGCGTCTCGTCTCCGTTGACTGCAAATGTTCAAGGGTACCAACGCTTCTTTGAGGAATAGA
GATGAAGAACTGCAGACAAGCAGATACAATTCAATGACCAAAGTTTGGGGGAAATGAC
TATGCACCCAAGGTACGGAAGCCATACAGATAACAAAAGAGAGAGAGATGGACAGAT
GAAGAGCACAAGAAGTTTGTGAAGCCTTGAAATTATACGGGCGAGCTTGGAGACGAATA
GAAGAACATGTGGGCTCAAAGACCGCAGTTCAGATTGCAAGCCATGCTCAGAAGTTTTC
TCTAAGGTTGCTCGAGAAGCAACTGGAGGTGATGGGAGCTCAGTAGAGCCGATTGTAATA
CCTCCTCCTCGTCCCAAGAGAAAGCCAGCGCATCCGTACCCTCGTAAGTTTGGGAACGAG
GCAGATCAAACAAGTAGATCGGTTTCTCCCTCAGAACGTGATACTCAATCTCCAACCTCT
GTGTTGTCCACTGTTGGATCAGAAGCATTGTGTTCCCTTGATTTCGAGTTACCCCAATCGA
AGCTTGTCCCCAGTTTCTTCTGCATCACCACCAGCTGCTCTTACAACCACTGCAAATGCA
CCTGAAGAGCTTGAGACTCTGAAGCTGGAGTTGTTTCTAGTGAGAGACTCTTAAACAGG
GAGAGCTCGATCAAGGAACCAACGAAGCAAAGTCTTAAACTCTTTGGGAAGACAGTTTG
GTATCTGATTACGGCATGTCTCTCTCTAACAACCTTCAACATATTGTAAATCCCCAATT
CAGCCATTACCACGGAACTCTCATCATCCAAGACACTACCCATAATAAGAACTCACAA
GAAGAAGCTCTTGAGCTGCTGGATACAAGTCCCTCTTAAGCAAGAAGATGTGGAAAATAGA
TGTTTGGATTACAGGAAAGCTGTCCAAAACGAAGGATCATCGACTGGATCAAACACTGGT
TCGGTGGATGATACGGGACACACGGAAAAGACCACAGAACCCGAAACAATGCTATGTCAA
TGGGAGTTTAAACCAAGTGAGAGGTCTGCATTTTCTGAGCTCAGAAGAACAACCTCCGAG
TCAAATTCAGAGGATTGGTCCATACAAGAAGAGAAAGATGGTAACAGAAGAAGAAGAG
CATGAGATTCTCTCCACTTATAA

>G1635 Amino Acid Sequence (domain in AA coordinates: 44-104)
MASSPLTANVQGTNASLRNRDEETADKQIQFNDQSFSGGNDYAPKVRKPYTITKERERWTD
EEHKKFVEALKLYGRAWRRIEHHVGSKTAVQIRSHAQKF FSKVAREATGGDGSSVEPIVI
PPRPKRKPAHPYPKRFNGEADQTSRSVSPSERDTQSPTSVLSTVGSEALCSLDSSSPNR
SLSPVSSASPPAALTTTANAPEELETLELFP SERLLNRESSI KEPTKQSLKLFKTVL

VSDSGMSSSLTTSTYCKSPIQPLPRKLSSSKTLPRIIRNSQEELLSCWIVPLKQEDVENR
CLDSGKAVQNEGSSTGSNTGSVDDTGHTEKTEPETMLCQWEFKPSERSAFSELRRTNSE
SNSRGFGPYKKRKMVTEBBEHEIHLHL*

>G1794 (160..1335)

TCTTTCTTTCTTCTCTTTGTCTCTGTTTCTGTTTCTCTCTCTCTCTCTCTACAGAGTT
TTCTTTCCCTCGAAGAAAAAGAAATATTTTTAAATTTAATTTTCTCTGCGTTTATAAGCTT
TAAGTTTCAGAGGAGGAGGATTTAGAAGGAGGGTTTTGTATGTGTGTCTTAAAAGTGGCA
AATCAGGAAGATAACGTTGGCAAAAAGCCGAGTCTATTAGAGACGATGATCATCGGACG
TTATCTGAAATCGATCAATGGCTTTACTTATTTCGAGCCGAAGACGACCACCACCGTCAT
AGCTTCCCTACGCAGCAGCCGCTCCATCGTCGTCGTCCTCATCTCTTATCTCAGGTTTC
AGTAGAGAGATGGAGATGTCTGCTATTGTCTCTGCTTTGACTCACGTTGTTGCTGGAAAT
GTTCTCTCAGCATCAACAAGGAGGCGGTGAAGGTAGCGGAGAAGGGACTTCGAATTCGTCT
TCTTCTCGGGGCGAGAAAAGGAGGAGAGAGGTGGAGGAAGGTGGCGCCAAAGCGGTTAAG
GCAGCTAATACTTTGACGGTTGATCAATATTTCTCCGGTGGTAGCTCTACTTCTAAAGTG
AGAGAAGCTTCGAGTAACATGTCAGGTCCGGGCCCAACATACGAGTATACAACCTACGGCA
ACTGCTAGTAGCGAAACGTCGTCGTTTAGTGGGGACCAACCCTCGGCGAAGATACAGAGGA
GTTAGACAAAGACCATGGGGAAAAGTGGGCGGCTGAGATTGAGATCCATTTAAAGCAGCT
AGAGTTTGGCTCGGTACGTTTCGACAATGCTGAATCAGCAGCAAGAGCTTACGACGAAGCT
GCACTTCGGTTTAGAGGCAACAAAGCCAACTCAACTTCCCTGAAAACGTCAAAACCTCGTT
AGACCTGCTTCAACCGAAGCACAACCTGTGCACCAACCGCTGCTCAAAGACCGACCCAG
TCAAGGAACTCGGGTTCAACGACTACCCTTTGGCCATAAGACCTGCTTCAATCAAAGC
GTTCAATTCGCAGCCGTTGATGCAATCATACAACCTTGAGTTACTCTGAAATGGCTCGTCAA
CAACAACAGTTTCAGCAACATCATCAACAATCTTTGGATTTATACGATCAAATGTCGTTT
CCGTTGCGTTTTCGGTCACACTGGAGGTTCAATGATGCAATCTACGTCGTCATCATCTCT
CATCTCTGTCCTCTGTTTTCCCGGGCTGCTGTTTCAGCCGCCACCAGAATCAGCTAGCGAA
ACCGGTTATCTCCAGGATATACAATGGCCATCAGACAAGACTAGTAATAACTACAATAAT
AGTCCATCCTCCTGATGACTTGCTTCATTTTATTTGTTTCACTATAGAGTAATAGAAAAC
AGGAAAATGATTATATGTTATAGAGTTATTTTCCAAATATTATAGGGTTTAGGTTGTTT
GTATTGTTCTGCTTTCATCCTCTCATGCTTTTTTTCTTAATTTATTATATTTTTGCATTA
TAATTTCTGTTTCATTGTAAACAAACATTAAAAAGACCACATGGAGAAAGGAAAAAAGAG
AG

>G1794 Amino Acid Sequence (domain in AA coordinates: TBD)

MCVLKVANQEDNVGKKAESIRDDDHRTLSEIDQWLYLFAAEDDHHRHSFPTQQPPSSSS
SSLISGFSREMEMSAIVSALTHVVGAVNPQHQQGGEGSGEGTSNSSSSSGQKRRREVEE
GGAKAVKAANTLTVDQYFSGSSSTSKVREASSNMSGPGPTYEYTTTATASSETSSFSGDQ
PRRRYRGVRQRPWGKWAABIRDPFKAARVWLGTFDNAESAARAYDEAALRFRGNKAKLNF
PENVKLVLPASTEAPVHQTAQRPTQSRNSGSTTTLLPIRPASNQSVHSQPLMQSYNLS
YSEMARQQQFQHQHQQSLDLYDQMSFPLRFGHTGGSMMQSTSSSSSHSRPLFSPAAPVQP
PPESASETGYLQDIQWPSDKTSNNYNSPSS*

>G1839 (38..592)

ATCACAGTTATGTTTCCATTCATTGGCTATAAAAAACCATGCTCACTCCCTTTTGTCTTC
ACACCATTTCAGGAAAAAATGAATAGTTGTCAGTCTAATCCCACCAAAATGGATAATTC
AGAAAATGTTCTATTTAATGATCAAAACGAAAATTTACACTTGTTCACCAACACCCCTTC
TTCTTCGTACTTTGACAAGAGATCAAGAGCACGAGATCATGGTCTCTGCTCTGCGACAAGT
GATATCTAACTCCGGAGCTGACGACGCGTCATCATCAAACCTTGATCATCAAGCGTTCC
GCCTCCAGACGCTGGCCCTTGTCCTCTCTGTGGCGTCGCCGTTGCTACGGCTGCACATT
ACAACGGCCGACCGAGAGGTAAAGAAGGAGAAGAAATACAAAGGAGTAAGGAAAAAACC
ATCGGGTAAATGGGCGGCGGAGATATGGGATCCGAGATCGAAATCAAGGAGGTGGCTTGG
AACGTTTCTTACGGCGGAGATGGCGGCACAATCTTACAATGATGCGGCGGCTGAGTATCG
AGCAAGACGTGGTAAACAAACGGAGAAGGAATTAACGGCGGTGGAGATGACTGAGAAG
GACATGGTCGGTGATCATACAGGCGAGGTGGAAATGTTATTTACTATTGAAAACATAA
ATTATTTATTATAGAGGAGATATACTCTTACGCTTTTCATTAAGATTTATTTTATAA
GTTTTAAAGTATTTTTATTGTTATAAAAAAAAAAAAAAAAAAAAAA

>G1839 Amino Acid Sequence (domain in AA coordinates: TBD)

MLTFFCSSHHLQEKMNCSQSNPTKMDNSENVLFNDQENFTLVAPHPSSSYLTRDQEHEI
MVSALRQVISNSGADASSSNLIITSVPPPDAGPCPLCGVAGCYGCTLQRPHEVKKKK
YKGVRRKPSGKWAABEWDPRSKSRRWLGTFLTAEMAAQSYNDAAEYRARRRGKTNBEGIK

RRWR*

>G2108 (35..694)

GAGAGAGAAACATTGATCTCTGAATATTGTGAACATGTTGAAATCAAGTAACAAGAGAAA
 AAGCAAAGAAGAGAAGAGTTACAAGAAGGGAAGTACCTTGGAGTGAGGAGACGTCCATG
 GGAAGATATGCAGCTGAAATCAGAAACCCCTTTTACTAAAGAAAGACATTGGCTTGAAC
 GTTTGATACAGCCGAAGAAGCTGCTTTTGCATATGACGTTGCTGCTCGATCCATCAGCGG
 CTCTCTAGCTACAACAACTTCTTCTACACTGAAAACACCTCTTTAGAAAGACATCCACA
 ACAGTCTTTGGAGCCTCATATGACTTGGGGATCTTCTAGTCTCTGTCTTCTTCAAGATCA
 GCCTTTTGAACAACCATTTTGTGCTGATCCTATCTCTTCTTCTTTTCTCAAAAACA
 AGAGTCTTCTACCAATCTCACTAACACTTTCTCACATTGTTATAATGATGGTGATCATGT
 TGGCCAAAGCAAGAGATTCTTTTACCTAATGATATGTCAAACAGTTTATTCCGTCATCA
 GGACAAAGTCGGTGAACATGACAATGCAGACCATAGAAGTTTGGCTCAGTTCTCAGCGA
 CGAACCTCTCTGCTTTTGAGTATGACTACATTGGGAATTATCTTCAAGAGTTTCTCAAAGA
 TGTCACGACGATGCTCCACAGTTTCTTATGTGAGCTTGTATTACCGATCCTTCAATTTA
 TG

>G2108 Amino Acid Sequence (domain in AA coordinates: 18-85)

MLKSSNKRKSKKEKKLQEGKYLGVRRRPWGRYAABIRNPFYTKERHWLGTDFDTEEAFAFAY
 DVAARSISGSLATTNFFYTENTSLEHRPQQSLEPHMTWGSSSLCLLQDQPFENNHFVADP
 ISSSFSQKQESSTNLNTNFSHCYNDGDHVGQSKEISLPNDMSNSLFGHQDKVGEHDNADH
 MKFGSVLSDEPLCFEYDYIGNYLSFQKDVNDAPQFLM*

>G2291 (27..797)

GCTTTCTACCTTTATAAAATAGAAAATGGAAAACAGCTACACCGTTGATGGTCACCGTC
 TTCAATATTCCGTTCCGTTAAGCTCCATGCATGAAACCAGTCAAACTCCGAACTTACG
 GATTATCCAAAGAGTCGCCGTTGGTCTGCATGCCCTTGTTGAAACCAACACTACTTCAT
 TCGATATCTCTTCTCTTTTCTCGTTTAAACCCAAAACAGAACCCGAAAACACGCATCGTG
 TCATGGACGATTCCATCGCCGCCGTCGTGGGCGAAAACGTTCTTTTCGGTGATAAAAAACA
 AAGTCTCTGATCACTTGACCAAAGAAGGTGGTGTGAAGCGGGGCGGAAGATGCCGCGAGA
 AGACCGGAGGATTTCATGGGAGTGAGAAAACGCCCGTGGGGGAGATGGTCGGCGGAGATAA
 GAGACAGGATAGGGCGGTGCAGACACTGGTTAGGAACGTTTCGACACGGCGGAAGAGGCAG
 CGCGTGCGTATGACGCGGCGGCGAGGAGGCTTAGAGGGACCAAAGCCAAGACCAATTTTCG
 TGATTCTCCGCTTTTTTCCCAAGGAAATAGCTCAGGCTCAGGAGGATAATAGGATGAGGC
 AGAAGCAGAAGAAGAAGAAGAAAAAGTGAGTGTGAGGAAGTGTGTTAAAGTCACAT
 CGGTTGCACAGTTGTTCGATGATGCCAATTTTATAAATTCTTCTAGTATTAAAGGAAATG
 TGATTAGTTCTATTGATAATCTTGAAAAATGGGTCTAGAGCTTGATTGTAGTTTAGGGT
 TGTGTCTAGGAAGTGATAAAGCACTCGTAGTTAAGTAGTTGTAGTT

>G2291 Amino Acid Sequence (domain in AA coordinates: TBD)

MENSYTVDGHLRLQYSVPLSSMHETSQNSETYGLSKESPLVCMPLFETNTTSFDISSLFSF
 NPKPEPENTHRVMDDSIAAVVGENVLFQDKNKSVDHLTKEGGVKRGKMPQKTGGFMGVR
 KRPWGRWSAEIRDRIGRCRHWLGTDFDTEEAARAYDAAARRLRGTAKTNFVIPPFPKE
 IAQAQEDNRMQRQKQKKKKKVSVRKCVKTSVAQLFDDANFINSSSIKGNVISSIDNLE
 KMGLELDLSLGLLSRK*

>G2452 (1..804)

ATGTCATCGTCGACGATGTACAGAGGAGTTAATATGTTTTACCGGCAAACACAACTGG
 ATTTTTCAAGAAGTCAGAGAAGCCACGTGGACGGCGGAGGAGAACAACGGTTCGAGAAA
 GCTCTCGCTTATCTGGACGACAAAGACAATCTTGAGAGCTGGTCCAAGATCGCAGATTG
 ATTCCCGGCAAACAGTAGCTGACGTCATTAAACGATACAAGGAGCTAGAGGATGATGTC
 AGCGACATCGAAGCCGACTTATCCCCATTCGGGATACGGCGGCGACGCCTCTCCGCT
 GCAAACAGTGACTATTTCTTTGGTCTAGAAAACCTCCAGCTACGGTTATGATTACGTCGTT
 GGAGGAAAGAGGAGTTCCGCCGCGATGACTGATTGTTTTAGGTCTCCGATGCCGAAAAG
 GAGAGGAAGAAAGGAGTTCCGTGGACCGAGGACGAACACCTACGATTCTGATGGGTTTG
 AAGAAATATGGAAAAGGAGATTGGAGAAACATAGCAAAAAGCTTTGTGACGACTCGAACG
 CCGACGCAAGTCGCTTCACACGCTCAGAAATATTTCTTCGACAACTCACAGATGGTAAA
 GACAAAAGACGATCAAGTATTCAGATATCACCAGTGTAAACATCCCTGACGCAGACGCA
 TCCGCAACCGCCACGACCGCTGACGTAGCACTCTCTCTACTCCAGCCAATTCTTTTGAC
 GTTTTCCTTCAGCCAAATCCTCATTACAGTTTCGCGTCTGCGTCTGCGTCTAGCTATTAT
 AATGCGTTTCCGCGAGTGGAGTTAA

>G2452 Amino Acid Sequence (conserved domain in AA coordinates: 27-213)

MSSSTMYRGVNMFPANTNWIQFQEVREATWTAENKRFKALAYLDDKDNLESWSKIADL
IPGKTVADVIKRYKELEDDVSDIEAGLIPIPGYGGDASSAANSYFFGLENSSYGYDYVV
GGKRSSPAMTDCFRSPMPEKERKKGVPWTEDEHLRFLMGLKKYKGDWRNIAKSFVTTRT
PTQVASHAQKYFLRQLTDGKDKRRSSIHDITTVNIPDADASATATTADVALSPTPANSFD
VFLQPNPHYSFASASASSYNAFPQWS*

>G2509 (143..934)

ATATATTCCCTCTTTCAATTCTCCTTCTTCGTCTTTTCTTTGTTTCTCATATTCAAGACAT
CCTCAATTCCAAATCTTAAACCCCTAAATTTACAGACACAATCGAGATCACCTGAAAAAAG
AGGTTTAAAGATTTTAGCAAAGATGGCGAATTGAGGAAATTTATGGAAAGAGGCCCTTTTCG
AGGCGATGAATCGGATGAAAAGAAAGAAGCCGATGATGATGAGAACATATCCCTTTCTT
CTCTGCCCGATCCCAATATGACATGCGTGCCATGGTCTCAGCCTTGACTCAAGTCATTGG
AAACCAAAGCAGCTCTCATGATAATAACCAACATCAACCTGTTGTGTATAATCAACAAGA
TCCTAACCCACCGGCTCCTCAACTCAAGATCAAGGGCTATTGAGGAAGAGGCACTATAG
AGGGGTAAGACAAAGCAGCATGGGGAAAGTGGGCAGCTGAAATTCGGGATCCGCAAAAGGC
AGCAGGGGTGTGGCTCGGGACATTTGAGACTGCTGAAGCTGCGGCTTTAGCTTATGATAA
CGCAGCTCTTAAGTTCAAAGGAAGCAAAGCCAACTCAATTTCCCTGAGAGAGCTCAACT
AGCAAGTAACACTAGTACAACCTACCGGTCCACCAAACTATTATTCTTCTAATAATCAAAT
TTACTACTCAAATCCGCAGACTAATCCGCAAAACCATACCTTATTTTAACCAATACTACTA
TAACCAATATCTTTCATCAAGGGGGGAATAGTAACGATGCATTAAAGTTATAGCTTGGCCGG
TGGAGAAACCGGAGGCTCAATGTATAATCATCAGACGTTATCTACTACAAATTCTTCATC
TTCTGGTGGATCTTCAAGGCAACAAGATGATGAACAAGATTACGCCAGATATTTGCGTTT
TGGGGATTCTTCACTCCTAATTCTGGTTTTGTAGATCTTCAATAAACTGATAATAAAGG
ATTTGGGTCACCTGTTATGAGGGGATCATATGTTTCTAA

>G2509 Amino Acid Sequence (domain in aa coordinates: 89-156)

MANSGNYGKRPFGRGDESEKKEADDDENIFPFPSARSQYDMRAMVSALTQVIGNQSSSHD
NNQHQPVVYNQDDPNPPAPPTQDQGLLRKRHYRGVVRQRPWGKWAABIRDPQKAARVWLGT
FETAEEAALAYDNAALKFKGSKAKLNFPERAQLASNTSTTTGPPNYSSNNQIYYSNPQT
NPQTIPIFYNQYYNYLHQGGNSNDALSYSLAGGETGGSMTNHQTLSTNNSSSSGGSSRQ
QDDEQDYARYLRFGDSSPPNSGF*

>G390 (1..2526)

ATGATGGCTCATCACTCCATGGACGATAGAGACTCTCCTGATAAAGGATTTGATTCCGGC
AAGTACGTTAGATACACGCCGAACAAGTTGAAGCTCTTGAGAGAGTTTATGCTGAGTGT
CCTAAACCTAGCTCTCTGAGAAGACAACAGCTTATTTCGTGAATGTCCCATTCTCTGTAAC
ATCGAGCCTCGACAGATCAAAGTTTGGTTCCAGAATCGCAGATGTCGAGAGAAGCAGAGG
AAAGAGTCAGCTCGTCTTCAGACAGTGAACAGGAAGCTGAGTGCTATGAACAAGCTTTTG
ATGGAAGAGAATGATCGTTTGCAAGCAAGTCTCCAACCTGGTTTATGAGAATGGATTCT
ATGAAACATCGAATCCACACTGCTTCTGGGACGACCACAGACAACAGCTGTGAGTCTGTG
GTCTGTGAGTGGTCAGCAACGTCAGCAGCAAAACCAACACATCAGCATCCTCAGCGTGAT
GTTAACAACCCAGCTAATCTTCTCTCGATTGCGGAGGAGACCTTGGCGGAGTTCCTTTGC
AAGGCTACAGGAAGTGTGTCGACTGGGTCCAGATGATTGGGATGAAGCCTGGTCCGGAT
TCTATTGGTATCGTAGCTGTTTCACGCAACTGCAGTGGAATAGCAGCACGTGCCTGTGGC
CTCGTGAGTTTGAACCCATGAAGGTGCTGAAATCCTCAAAGATCGTCCATCTTGGTTC
CGTGACTGTGATGTGTCGAGACTCTGAATGTTATACCACTGGAAATGGTGGTACTATC
GAGCTTGTCAACACTCAGATTTATGCTCCTACAACATTAGCAGCAGCTCGTGACTTTTGG
ACGCTGAGATATAGTACAAGTCTAGAAGATGGAAGCTATGTGGTCTGTGAGAGATCACTC
ACTTCTGCAACTGGTGGCCCCAATGGTCCACTTTCTTCAAGCTTCGTGAGAGCCAAAATG
CTGTCAAGCGGGTTTCTTATCCGTCCTTGTGATGGTGGTGGTTCCATTATTCATCGTT
GATCATGTGGACTTGGATGTCTCAAGTGTTCCCTGAAGTCCTCAGGCCTCTTTATGAGTCT
TCCAAAATCCTTGCTCAAAAAATGACTGTGCTGCTCTGAGACATGTGCGCCAAAATGCT
CAAGAGACTAGTGGAGAAGTCCAGTATAGTGGTGACGCCAGCCTGCAGTTTAAAGACT
TTCAGCCAGAGACTCTGCCGGGGTTTCAATGATGCTGTAAATGGTTTTGTGATGATGGA
TGGTCTCCAATGAGTAGTGATGGAGGAGGATATTACGATCATGATTAACTCTTCCTCT
GCTAAATTTGCTGGCTCCCAATACGGTAGCTCATTTCTTCCAAGTTTTGGAAGTGGTGT
CTCTGTGCCAAAGCTTCTATGCTGTTGCAGAATGTTCCACCCCTTGATTGATTGGGTTT
CTGAGAGAACACCGAGCTGAATGGGCAGACTATGGTGTGATGCCTATTCTGCTGCATCT
CTCAGAGCAACTCCATATGCTGTTCCATGCGTCAGAACCGGTGGGTTCCCGAGTAACCAA
GTCATTCTTCTCTCGCACAGACACTCGAACATGAAGAGTTTCTCGAAGTGGTTAGACTT

GGAGGTCATGCTTACTCACCTGAAGACATGGGCTTATCCCGGGATATGTATTTACTGCAG
CTTTGTAGCGGCGTTGATGAAAAATGTGGTTGGAGGTTGTGCTCAGCTTGTCTTTGCCCA
ATCGATGAATCATTTGCTGATGATGCACCTTTGCTTCTCTTCTGGTTTCCGTGTCATACCA
CTCGACCAAAAAACAAATCCGAATGATCATCAATCTGCAAGTCGAACACGGGATCTAGCA
TCGTCCCTAGATGGTTCCACCAAAACCGATTCCGAAAACAACTCTAGATTGGTCTTAACA
ATAGCCTTCCAGTTCACGTTTGATAACCATTCCAGAGACAATGTTGCTACAATGGCGAGA
CAGTATGTGAGGAACGTTGTTGGTTCGATTAGAGAGTGGCTCTAGCCATTACGCCTCGT
CCTGGCTCAATGCAACTTCCCACTTCCCTGAAGCTCTCACTCTTGTCCGTGGATCACC
CGTAGTTACAGTATTCATACAGGTGCAGATCTGTTTGGAGCTGATTCTCAGTCCTGTGGA
GGAGACACATTGCTTAAGCAACTCTGGGACCATAGTGATGCCATATTGTGCTGCTCCCTG
AAAATAATGCCTCACCAGTATTACATTTGCAAACCAAGCTGGTTTAGACATGCTTGAA
ACTACACTTGTGGCACTTCAGGATATAATGCTCGACAAAACACTTGATGACTCTGGTCGT
AGAGACTTTTGTGCTCCGATTCCGCAAGATCATGCAGCAGGGATATGCGAATCTTCCGGCA
GGAATATGTGTGTCGAGCATGGGCAGACCGGTTTCTGATGAGCAAGCGACGGTGTGGA
GTTGTTGATGACAACGAATCAAACCACTGCTTGGCTTTTACCCTCGTTAGTTGGTCGTTT
GTTTGA

>G390 Amino Acid Sequence (domain in AA coordinates: 18-81)
MMAHHSMDDRSPDKGFDGSKYVRYTPEQVEALERVYAECPKPSSLRRQQLIRECPILCN
IEPRQIKVWFQNRRCREKQRKESARLQTVNRKLSAMNKLMEENDRLQKQVSNLVYENG
MKHRIHTASGTTTNSCESVSVSGQQRQQNPTHQHPQRDVNNPANLLSIAETLAELFC
KATGTAVDWVQMIGMKPGPDSIGIVAVSRNCSGIAARACGLVLEPMKVABEILKDRPSWF
RDCRCVETLNVIPITGNGGTIELVNTQIYAPTTLAAARDFWTLRYSTSLLEDGSYVVCERSL
TSATGGPNGPLSSSFVRKMLSSGFLIRPCDGGGSI IHIVDHVDLDVSSVPEVLRPLYES
SKILAQMKTVAALRHVRQIAQETSSEVQYSGGRQPAVLRTFSQRLCRGFNDVNGFVDDG
WSPMSSDGGEDITIMINSSSAKFAGSQYSSFLPSFGSGVLCAKASMLLQNVPLVLIRF
LREHRAEWADYGVDAYSASLRATPYAVPCVRTGGFPSPNQVILPLAQTLHEEFLEVRL
GGHAYSPEDMGLSRDMYLLQLCSGVDEENVVGGCAQLVFAPIDESFADDAPLLPSGFRVIP
LDQKTNPNDHQASRTRDLASSLDGSTKTDSETNSRLVLTIAFQFTFDNHSRDNVATMAR
QYVRNVVGSIQRVLAITPRPGSMQLPTSPREALTLVRWITRSYSIHTGADLFGADSQSCG
GDTLLKQLWDHSDAILCCSLKTNASPVFTFANQAGLDMLETTLVALQDIMLDKTLDDSGR
RALCSEFAKIMQQGYANLPAGICVSSMGRPVSYEQATVWKVDDNESNHCLAFTLVSWSF
V*

>G391 (1..2559)
ATGATGATGGTCCATTTCGATGAGCAGAGATATGATGAACAGAGAGTCGCCGGATAAAGGG
TTAGATTCGGCAAGTATGTGAGGTACACGCCGAGCAAGTGGAAGCTCTCGAGAGAGTT
TACACTGAGTGCTCTAAGCCAAGTTCTCTAAGAAGACAACACTCATACGTGAATGTCCG
ATTCTCTTAACATCGAGCCTAAGCAGATCAAAGTTTGGTTTCAGAACCAGCAGATGTCGT
GAGAAGCAGAGGAAAGAAGCTGCTCGTCTTCAAACAGTGAACAGAAAACCTCAATGCCATG
AACAACTCTTGATGGAAGAGAATGATCGTTTGCAGAAGCAAGTTTCTAACTTGGTCTAT
GAGAATGGCCACATGAAACATCAACTTCACACTGCTTCTGGGACGACCACAGACAACAGC
TGTGAGTCTGTGGTTCGTGAGTGGTCAGCAACATCAACAGCAAAACCCAAATCCTCAGCAT
CAGCAACGTGATGCTAACAACCCAGCAGGACTCCTTCTATAGCAGAGGAGGCCCTAGCA
GAGTTCCTTTTCAAGGCTACAGGAAGTGTGTTGACTGGGTTCAGATGATTGGGATGAAG
CCTGGTCCGGATTCTATTGGCATAGTCGCTATTTTCGCGCAACTGCAGCGGAATTGCAGCA
CGTGCCCTGCGCCCTCGTGAGTTTGAAGCCATGAAGGTTGCTGAAATTCTCAAAGATCGT
CCATCTTGGCTCCGAGATTGTGCAAGTGTGGATACTCTGAGTGTGATACCTGCTGGAAAC
GGTGGGACGATCGAGCTTATTTACACGCAGATGTATGCTCCTACGACTTTAGCAGCAGCT
CGTGACTTTTGGACGCTGAGATATAGCACATGTTTGAAGATGGAAGCTATGTGGTTTGT
GAAAGGTCGCTTACTTCTGCAACTGGTGGCCCCACTGGGCCACCTTCTTCAAACCTTTGTG
AGAGCTGAAATGAAACCAAGCGGGTTTCTCATCCGTCTTGCGATGGTGGTGGTTCATT
CTCCACATTGTTGATCATGTTGATCTGGATGCCTGGAGTGTCCCTGAAGTCATGAGGCCT
CTCTATGAATCATCGAAGATTCTTGCTCAGAAAATGACTGTTGCTGCTTTGAGACATGTA
AGACAAATTGCACAAGAAACAAGTGAGAAAGTTCAAGTATGGTGGAGGGCGCCAACCTGCG
GTTTTAAGAACCTTCAGTCAAAGACTCTGTGGGGTTTCAATGATGCTGTTAATGGTTTTT
GTGGATGATGGATGGTCACCAATGGGTAGCGATGGTGCAGAGGATGTTACTGTAATGATA
AACTTGTCCCCTGGGAAGTTTGGTGGGTCTCAGTACGTAATTCATTCTTCCAAGCTTT
GGTAGTGGCGTGCTTTGTGCCAAGGCATCTATGTTGCTTCAGAACGTTCCACCCGCTGTG

CTGGTTTCGATTCCCTTAGAGAACACCGATCTGAATGGGCTGATTATGGCGTGGATGCTTAT
 GCTGCTGCATCGCTCAGAGCAAGTCCTTTTGTGCTTCTTGTGCTAGAGCTGGGGGGTTC
 CCAAGTAACCAAGTCATTCCTCTTTCGCGACAGAGTTGAACATGAAGAGTCACTTGAG
 GTGGTTAGACTTGAAGGTCACGCTTACTCACCCGAAGACATGGGTTTAGCTCGGGATATG
 TATTTGCTACAGCTTTGTAGCGGTGTTGATGAAAATGTGGTTGGAGTTGTGCACAGCTT
 GTATTTGCCCCATCGATGAATCATTTGCTGATGATGCACCTTTGCTTCCTTCCGGTTTC
 CGCATCATACCTCTTGAACAGAAATCTACTCCGAACGGTGCATCTGCAAACCGTACCCTG
 GATTTAGCCTCAGCTTTAGAAGGATCCACACGTCAGCTGGTGAAGCCGACCCAAATGGC
 TGTAACCTTAGGTCGGTACTAACCATAGCATTCCAGTTCACATTTGATAACCATTCAAGA
 GACAGTGTGCTTCAATGGCACGTCAGTACGTGCGAAGCATAGTAGGATCGATTAGAGG
 GTTGCTCTAGCCATTGCTCCTCGTCTGGCTCCAATATCAGTCCAATATCTGTTCCCACT
 TCCCCTGAAGCTCTCACTCTGGTCCGTGGATCTCCCGAGTTACAGCCTTCACACTGGT
 GCAGATCTCTTTGGATCTGATTCTCAAACAGTGGTGACACGTTGCTGTCATCAACTCTGG
 AATCACTCTGATGCAATCTTGTGCTGCTCCCTCAAAACAAACGCTTCACCGGTTTTCA
 TTCGAAACCAACCGGTTTAGACATGCTGGAACGACTCTTGTAGCCCTTCAAGACATA
 ATGCTAGACAAGACCTTGTGACGAACCTGGTCGTAAAGCTCTTGTCTGAGTTCCCAAG
 ATCATGCAACAGGGCTATGCTCATCTGCCGCGAGGAGTATGTGCGTCAAGCATGGGAAGG
 ATGGTATCTTACGAGCAGGCAACCGTGTGGAAAGTTCTTGAAGACGATGAATCAAACCA
 TGCTTAGCTTTCATGTTCTGTAATTGGTCTGTTCTGTTGA

>G391 Amino Acid Sequence (domain in AA coordinates: 25-85)
 MMMVHSMRDMNRESPTDKGLDSGYVRYTPEQVEALERVYTECPKPSLRRQQLIRECP
 ILSNIEPKQIKVWFQNRRCREKQRKEAARLQTVNRKLNAMNKLMEENDRLQKQVSNLVY
 ENGHMKHQLHTASGTTTNDNSCESVVVSGQQHQQQNPNPQHQQORDANNPAGLLSIAEELA
 EFLSKATGTAVDWVQMIGMKPGPDSIGIVAISRNCSGIAARACGLVLEPMKVABILKDR
 PSWLRDCRSVDLTSLVIPAGNGGTIELIYTQMYAPTLLAAARDFWTLRYSTCLEDSYVVC
 ERSLSATGGPTGPPSSNFVRAEMKPSGFLIRPCDGGGSILHIVDHDLDAWSVPEVMRP
 LYESSKILAQKMTVAALRHVRQIAQETSSEVQYGGGRQPAVLRTFSQRLCRGFNDVNGF
 VDDGWSPMGSDGAEDVTVMINLSPGKFGGSQYGNFLPSFGSGVLCAKASMLLQNVPPAV
 LVRFLREHRSEWADYGVDAAYAAASLRASPFVPCARAGGFPSNQVILPLAQTVHEESLE
 VVRLEGHAYSPEDMGLARDMYLLQLCSGVDEENVVGGCAQLVFAPIDESFADDAPLLPSGF
 RIIPLEQKSTPNGASANRTLDLASALEGSTRQAGEADPNCNFRSVLTIAFQFTFDNHSR
 DSVASMARQYVRSIVGSIQRVALAIAPRPGSNISPISVPTSPEALTLVRWISRSYSLHTG
 ADLFGSDSQTSQDITLLHQLWNHSDAILCCSLKTNASPVFTFANQTGLDMLETTLVALQDI
 MLDKTLDEPGRKALCFEPKIMQQGYAHLPAVCASSMGRMVSYEQATVWKVLEDDSNH
 CLAFMFVNWSFV*

>G438 (188..2716)
 CGGGGTACCAAGCCACGACCGTAGAATCTTCTTTTGTCTGAAAAGAATTACAATTTACG
 TTTCTCTTACGATACGACGGACTTTCCGAAGAAATTAATTTAAAGAGAAAAGAAGAA
 GCCAAAGAAGAAGAAGAAGCTAGAAGAAACAGTAAAGTTTGAGACTTTTTTTGAGGGTCG
 AGCTAAAATGGAGATGGCGGTGGCTAACCAACCGTGAGAGAAGCAGTGACAGTATGAATAG
 ACATTTAGATAGTAGCGGTAAGTACGTTAGGTACACAGCTGAGCAAGTCGAGGCTCTTGA
 GCGTGTCTACGCTGAGTGTCTTAAGCCTAGCTCTCTCCGTCGACAACAATTGATCCGTGA
 ATGTTCCATTTTGGCCAATATTGAGCCTAAGCAGATCAAAGTCTGGTTTCAGAACCGCAG
 GTGTCGAGATAAGCAGAGGAAAGAGGCGTCGAGGCTCCAGAGCGTAAACCGGAAGCTCTC
 TGCGATGAATAAACTGTTGATGGAGGAGAATGATAGGTTGCAGAAGCAGGTTTCTCAGCT
 TGTCTGCGAAAATGGATATATGAAACAGCAGCTAACTACTGTTGTTAACGATCCAAGCTG
 TGAATCTGTGGTCACAACTCCTCAGCATTCGCTTAGAGATGCGAATAGTCTGCTGGATT
 GCTCTCAATCGCAGAGGAGACTTTGGCAGAGTTCCTATCCAAGGCTACAGGAAGTCTGT
 TGATTGGGTTTCAGATGCCTGGGATGAAGCCTGGTCCGGATTCCGGTTGGCATCTTTGCCAT
 TTCGAAAAGATGCAATGGAGTGGCAGCTCGAGCCTGTGGTCTTGTAGCTTAGAACCTAT
 GAAGATTGCAGAGATCCTCAAAGATCGGCCATCTTGGTTCCGTGACTGTAGGAGCCTTGA
 AGTTTTCACTATGTTCCCGGCTGGTAATGGTGGCACAATCGAGCTTGTATATGCAGAC
 GTATGCACCAACGACTCTGGCTCTGCGCGGATTTCTGGACCCTGAGATACACAACGAG
 CCTCGACAATGGGAGTTTTGTGGTTTGTGAGAGGTCGCTATCTGGCTCTGGAGCTGGGCC
 TAATGCTGCTTCAGCTTCTCAGTTTGTGAGAGCAGAAATGCTTTCTAGTGGGTATTTAAT
 AAGGCCTTGTGATGGTGGTGGTTCTATTATTACATTGTGATCACCTTAATCTTGAGGC
 TTGGAGTGTTCGGATGTGCTTCGACCCCTTTATGAGTCATCCAAAGTCGTTGCACAAAA

AATGACCATTTCCGCGTTGCGGTATATCAGGCAATTAGCCCAAGAGTCTAATGGTGAAGT
AGTGTATGGATTAGGAAGGCAGCCTGCTGTTCTTAGAACCTTTAGCCAAAGATTAAGCAG
GGGCTTCAATGATGCGGTTAATGGGTTTGGTGACGACGGGTGGTCTACGATGCATTGTGA
TGGAGCGGAAGATATTATCGTTGCTATTAACCTCTACAAAGCATTGGAATAATATTTCTAA
TTCTCTTTTCGTTCTCTGGAGGCGTGCTCTGTGCCAAGGCTTCAATGCTTCTCCAAAATGT
TCCTCCTGCGGTTTTGATCCGGTTCCTTAGAGAGCATCGATCTGAGTGGGCTGATTTCAA
TGTTGATGCATATTTCCGCTGCTACACTTAAAGCTGGTAGCTTTGCTTATCCGGGAATGAG
ACCAACAAGATTCAGTGGGAGTCAGATCATAATGCCACTAGGACATACAATTGAACACGA
AGAAATGCTAGAAGTTGTTAGACTGGAAGGTCATTCTTTGCTCAAGAAGATGCATTTAT
GTCACGGGATGTCCATCTCCTTCAGATTTGTACCGGATTTGACGAGAATGCCGTTGGAGC
TTGTTCTGAACTGATATTTGCTCCGATTAATGAGATGTTCCCGGATGATGCTCCACTTGT
TCCCTCTGGATTCCGAGTCATACCCGTTGATGCTAAAACGGGAGATGTACAAGATCTGTT
AACCGCTAATCACCGTACACTAGACTTAACCTTAGCCTTGAAGTCGGTCCATCACCTGA
GAATGCTTCTGGAAACTCTTTTTCTAGCTCAAGCTCGAGATGTATTCTCACTATCGCGTT
TCAATTCCTTTTGAAAACAACCTGCAAGAAAATGTTGCTGGTATGGCTTGTCAGTATGT
GAGGAGCGTGATCTCATCAGTTCAACGTGTTGCAATGGCGATCTCACCGTCTGGGATAAG
CCCAGTCTGGGCTCCAAATTGTCCCAGGATCTCTGAAGCTGTTACTCTTGCTCAGTG
GATCTCTCAAAGTTACAGTCATCACTTAGGCTCGGAGTTGCTGACGATTGATTCACTTGG
AAGCGACGACTCGGTACTAAAACCTTATGGGATCACCAAGATGCCATCCTGTGTTGCTC
ATTAAAGCCACAGCCAGTGTTTATGTTTGCGAACCAAGCTGGTCTAGACATGCTAGAGAC
AACACTGTAGCCTTACAAGATATAACACTCGAAAAGATATTCGATGAATCGGGTCGTAA
GGCTATCTGTTCCGACTTTCGCCAAGCTAATGCAACAGGGATTTGCTTGCTTGCTTCCAGG
AATCTGTGTGTCACGATGGGAAGACATGTGAGTTATGAACAAGCTGTTGCTTGGAAGT
GTTTGCTGCATCTGAAGAAAACAACAACATCTGCATTGTCTTGCTTCTCCTTTGTAAA
CTGGTCTTTTGTGTGATTGATTGACAGAAAAGACTAATTTAAATTTACGTTAGAGAAC
TCAAATTTTGGTTGTTGTTTAGGTGCTCTGTTTTGTTTTTAAATTTATTTTGATCAA
A

>G438 Amino Acid Sequence (domain in AA coordinates: 22-85)
MEMAVANHRERSSDSMNRHLDSGKYVRYTAEQVEALERVYAECPKSSLRRQQLIRECS
ILANIEPKQIKVWFQNRRCRDKQRKEASRLQSVNRKLSAMNKLMEENDRLQKQVSQLVC
ENGYMKQQLTTVNDPSCSVVTTTQHSRLDANSPAGLLSIAEETLAEFLSKATGTAVDW
VQMPGMPKPGPDSVGIFAIQSRCNGVAARACGLVSLPEMKIAEILKDRPSWFRDCRSLEVF
TMFPAGNGGTIELVYMQTYAPTTLAPARDFWTLRYTSLDNQSFVVCERSLSGSGAGPNA
ASASQFVRAEMLSSGYLIRPCDGGGSI IHIVDHLNLEAWSVPDVLRLPLYESSKVVAQKMT
ISALRYIRQLAQESNGEVVYGLGRQPAVLRTFSQRLSRGFNDVNGFGDDGWSTMHCDGA
EDIIVAINSTKHLNNISNSLSFLGGVLCASMLLQNVPPAVLIRFLREHRSEWADFND
AYSAAITKAGSFAYNPMRPTFRFTGSQIIMPLGHTIEHEEMLEVVRLEGHSLAQEDAFMSR
DVHLLQICTGIDENAVGACSELI FAPINEMFPDDAPLVP SGFRVIPVDAKTGDVQDLLTA
NHRTLDLTSSLEVGPSPENASGNSFSSSSSRCILTIAFQFPFENNLOENVAGMACQYVRS
VISSVQVRVAMAI SPGISPSLGSKLSPGSPEAVTLAQWISQSYSHHLGSELLTIDSLGSD
DSVLKLLWDHQDAILCCSLKPQPVFMFANQAGLDMLETTLVALQDITLEKIFDESGRKAI
CSDFAKLMQGGFACLP SGICVSTMGRHVSYEQAVAWKVFAASEENNNLHCLAFS FVNWS
FV*

>G47 (38..472)

CTTCTTCTTCATCGATCATCATACAACAACAAAAATGGATTACAGAGAATCCACCGG
TGAAAGTCAGTCAAAGTACAAAGGAATCCGTCGTCGGAAATGGGGCAAATGGGTATCAGA
GATTAGAGTTCCGGGAACCTCGTGACCGTCTCTGTTTAGGTTTCACTCAACAGCAGAAGG
TGCCGCCGTAGCACACGACGTTGCTTTCTTCTGTTTACACCAACCTGATTCCTTAGAATC
TCTCAATTTCCCTCATTTGCTTAATCCTTCACTCGTTTCCAGAACTTCTCCGAGATCTAT
CCAGCAAGCTGCTTCTAACGCCGGCATGGCCATTGACGCCGGAATCGTCCACAGTACCAG
CGTGAACCTCTGGATGCGGAGATACGACGACGTATTACGAGAATGGAGCTGATCAAGTGGA
GCCGTTGAATATTTAGTGATGATTATCTGGGCGGCCACGATCACGTTTGATTTATCTC
GACGGTCATGATCACGTTTGATCTTCTTTGAGTAAGATTTGTACCATAATCAAAACAG
GTGTGGTGCTAAAAATCTTACTCAAAACAAGATTAGGTACCACAGAGAAACAATCAAATGG
TTGTGAATATACATTATAAGGTTTTGATTAATGTTTGTCTTCACTGATTTAGTGAAGTTTG
GTCCATTGTATACAAATCTATTCAAGAAACCTAGCGCGAGATCATGTTTCGTGATTGAAG
ATTGAGATTTTAAAGTATTCTGAATATTTTGTAAAAATACAAATAAAAAAAAAAAAAAA

AAAAA

>G47 Amino Acid Sequence (domain in AA coordinates: 11-80)

MDYRESTGESQSKYKGIRRRKWGKVVSEIRVPGTRDRLWLGSFSTAEGAAVAHDVAFFCL
HQPDSLESILNFPHLNPSLVSRTPRSIQQAASNAGMAIDAGIVHSTSVNSGCGDTTTY
ENGADQVEPLNISVYDYLGGHDHV*

>G559 (89..1285)

aaagttgctagctttaatttgccaacttactattcttatgtgtaataatcgtttgagg
tcgttgatttggtgataagtcagtagaaATGgataaggagaaatctccagcacctccttg
tggaggctcttcctcctccatctccatcaggtcgatgctctgcattctcagaagctggtcc
cattggtcatggttcagatgctaatacgaatgagtcgatattagccgtatgcttgataa
cccacctaagaagattggacatcggcgagctcattctgaaatacttactctccctgatga
tttgagctttgatagtgatcttggtgtggttggttaagtctgctgatggagcttctttctc
tgatgagactgaagaagatttgctctctatgtatcttgatattggataagtttaattcttc
tgctacatcttctgcccgaagttggtgagccatcaggaactgcttgaaaaatgagacaat
gatgcagacaggcacaggctcaacttccaatcctcagaatacggttaatagtcttgccga
aaggccaagaatcaggcatcaacatagccaatctatggatggttcaatgaatatcaatga
gatgcttatgtcgggaatgaagatgattctgctattgatgctaagaagtctatgtctgc
tactaaacttgctgagcttgctctcattgatcctaaacgtgctaaggagatattgggcaaa
caggcagtcgcgagcagcatcaaaagaaaggaagacgagatacatatttgagcttgagag
aaaagtacagactttgcaaacagaggctacaactctctcagcccagttgacctcttaca
gagagacacaaatggcttgactgttgaaaacaatgagctgaagctgcggttacaacaat
ggagcagcaggttcacttgccaggtgaactaaacgaagcactaaaggaggaaatccagca
tctgaaggtgttgactggccaagttgctccatcagcgttgaactatgggtcggttgatc
aaaccagcagcaattctattccaacaatcagtcattgcaacaatcttagctgcaaaaca
gttccagcaacttcagattcattcagaagcagcaacaacaacaacaacaacaaca
gcaacaccaacagcagcagcagcaacagcaacagtatcagtttcaacagcaacagatgca
acagcttatgcagcagcggcttcaacagcaagaacaacaaatggagtaagactcaagcc
ttcacaagcccagaaagagaacTGAggaatatgaatatgtcccacgtaagtgcagaggttc
tccttctgaacaatttccttctcattcataaattgttggttcattccatcacttgagcttc
ttggatttttagggtttttagctaacaca

>G559 Amino Acid Sequence (domain in AA coordinates: 203-264)

MDKEKSPAPPCGGLPPSPSGRCSAFSEAGPIGHGSDANRMSHDISRMLDNPPKKIGHRR
AHSEILTLPPDDLFDSDLGVVGNADGASFSDTEEDLLSMYLDMDKFNSSATSSAQVGE
PSGTAWKNETMMQTGTGSTSNPQNTVNSLGERPRIHQHSQSMDSMNINEMLMNSGNEED
SAIDAKKSMSATKLAELALIDPKRAKRIWANRQSAARSKERKTRYIFELERKVQTLQTEA
TTLAQQLTLQDRTNGLTVENNELKRLQTMQQVHLQDELNEALKEEIQHLKVLTGQVA
PSALNYGSFSGNQFYSNNQSMQTLIAAKQFQQLQIHSQKQQQQQQQQQQHQQQQQQQ
QQYQFQQQQMQQLMQRLQQQEQQNGVRLKPSQAQKEN*

>G568 (141..995)

GACCGGCTAAAGTCAAGAACCTCTCTCTGAGCTCTCACCACCTTCTCTCTACTCCCTC
TCTGCGTGTAGGATACTACTAGACAATTGACAACCAAAGACTAAAGCTGTGTTGTTGGTT
CACTTCTGTTCTCTTTTCCAATGTTGTCTATCAGCTAAGCATCAGAGAAACCATAGACTCT
CTGCTACAAACAAGAACCAGACTCTCACCAGTTTCTTCCATTTTCATCCTCATCACCAT
CGTCTTCTTCTTCATCATCATCAACCTCATCATCTCCTTTACCTTCTCAAGACTCTC
AAGCCCAAGAAGAGATCTCTTGTCAACATGGAAGAAGTTTGAATGACATCAACCTTGCTT
CCATCCACCACCTAAACCGACACAGCCCTCATCCACAACAACCACGAGCCAAGGTTCA
GGGGCCAAAACCACCACAACCAAAACCCTAACTCAATCTTCCAAGATTTTCTCAAAGGAT
CTTTGAACCAGGAAGCAGCACCCACAAGCCAGACCACGGGTTCTGCGCTAATGGCGATT
CCACCACGGTCACTGTTCTTTACAGCTCTCCTTTCCACCTCCTGCAACTGTTCTGAGCT
TGAATTCGGGCGCTGGCTTCGAGTTTCTCGATAACCAAGATCCTCTTGTACCTCAAACCT
CTAATCTTCATACCCACCATCACCTCTCAAACGCTCATGCCTTCAACACCTCTTTCGAGG
CTCTGGTTCATCCAGTTCTTTTGGTAAGAAAAGAGGCCAAGATTCCAATGAAGGTTTCAG
GGAATAGAAGACATAAGCGTATGATCAAGAACAGAGAATCTGCAGCTCGTTCCCGCGCTA
GGAAACAGGCTTATACAAACGAGTTAGAACTTGAAGTTGCTCACTTGCAGGCAGAAAATG
CAAGACTCAAGAGACAACAAGATCAAAAAATGGCTGCAGCAATTCAGCAACCAAAAAGA
ACACACTTCAACGGTCTTCCACAGCTCCATTTTGAGAAATCTACAAGTCTTGTCTCTCT
TTTGGGGATTGAGATTGTCTCATGAAGAAGTGAAAAATGGCAAAAGTTTGTACCCCTTTT

TTATTAGCTATAAGTATAACTAAGCCTAAAATTGTAGAACTAAGATATTGTAGGGGAAAA
AAGAAGATGTAAAACAAAAGACCCGGAAAGAGAAAAGGATCTTTCAATTTCTTAAGGCAC
AGGAACACCTGTCTGGGTCTCTCTTAATGTTCTGTCGTTTTCTTATGCAAACCCTTTT
TTCACCTCTGTACTAACTTATACTTGTATTCTTG

>G568 Amino Acid Sequence (domain in AA coordinates: 215-265)
MLSSAKHQHRLSATNKNQTLTKVSSISSSSPSSSSSSSTSSSSPLPSQDSQAQKRS
VTMEVWVDINLASIHHLNRHSPHPQHNHEPRFRGQNHNNQNPNSIFQDFLKGSNLNQEP
PTSQTTGSAPNGDSTTVTVLYSSPFPPPATVLSLNSGAGFEFLDNQDPLVTSNSNLHTHH
HLSNAHAFNTSFEALVPSSSFGKKRGQDSNEGSGNRRHKRMKKNRESAARSRRARKQAYTN
ELELEVAHLQAENARLKRQDQKMAAAIQPKKNTLQRSSTAPF*

>G580 (43..747)

CCAAAAACAAAGCATTCTATGCTATTCTGTTCTGTCTCCAATGTTGTCATCAGCAAAG
CATAATAAGATCAACAACCATAGTGCCTTTTCAATTTCTCTTCATCATCATCATTATCA
ACATCATCCTCCCTAGGCCATAACAAATCTCAAGTCACCATGGAAGAAGTATGGAAAGAA
ATCAACCTTGGTTCACTTCACTACCATCGGCAACTAAACATTGGTCATGAACCAATGTTA
AAGAACCAAAACCCTAATAACTCCATCTTTCAAGATTTCTCAACATGCCTCTGAATCAA
CCACCACCACCACCACCACCCTTCTCTTCCACCATTGTCACTGCTCTCTATGGCTCT
CTGCCTCTTCCGCCCTCTCGCCACTGTCTCAGCTTAAACTCCGGTGTGGATTGAGTTT
CTTGATACCAAGAAAATCTTCTGCTTCTAACCTCGCTCTTTGAGGAATCTGCAAAG
TTTGGTTGTCTTGGTAAGAAAAGAGGCCAAGATTCTGATGATACTAGAGGAGACAGAAGG
TATAAGCGTATGATCAAGAACAGAGAATCTGCTGCTCGTTCAAGGGCTAGGAAGCAGGCA
TATACAAACGAACCTTGAGCTTGAAATTGCTCACTTGCAGACAGAGAATGCAAGACTCAAG
ATACAACAAGAGCAGCTGAAAATAGCCGAAGCAACTCAAAACCAAGTAAAGAAAACACTA
CAACGGTCTTCCACAGCTCCATTTTGAGAAAAATCTACTATTTCTTTTGGGGGAGTTTC
AAGTGTCTCTTATGAAGATGAGAAAAACAGAAAAAGTTGTACATTTTAGCTAAGTTAAA
TTTGTGGTGGTAAGTAATGTAAGAAAAGTGTGTGTAGAAGAAAAGTGTCTAGAAAAAG
AAAGCAACTAACTTTCTTCTTCTCTCTGTTTCTTATCAACTCTTTTGACTTTTGTACT
TTTTTTCTTCTACTTAACCTCTATTATTGTAATGCCAAGTCAAGTCCTTATCTAGCTA
GTACATGAGTTTCTGTTTTCTGTTTAAAGCCAT

>G580 Amino Acid Sequence (domain in AA coordinates:162-218)
MLSSAKHNKINNHSAFSISSSSSSLSTSSSLGHNKSQVTMEEVWKEINLGSLSHYHRLNI
GHEPMLKNQNPNSIFQDFLNMPLNQPPPPPPSSSTIVTALYGSPLPLPPPATVLSLNS
GVGFEFLDTTENLLASNPRSFESAKFGCLGKKRGQSDSDTRGDRRYKRMKKNRESAARS
RARKQAYTNELELEIAHLQENARLKIQQEQLKIAEATQNVKKTQLQRSSTAPF*

>G615 (197..1252)

TTTTTTCTTTCTTTCTTTTCTTTTCTGCTGGTGTGAGAAATGTACGCTTACTATCTCTCTCT
CTCTCTGCCAGATTCTCTCTTTTGGATGATGTGAAAGTTGTGCTTTTGTCTTAAAGAAA
AAGGCATATTTTAAATACTTGATTCTTGGTCTTGATTCTTGATTCTTGGTTTTTTTAG
CTTCTTAAGTTCCGTGATGTCGCTCTCCACCAATGACTACAACGATGGTAATAACAATGG
AGTGTACCCTCTCTCTCTTTACCTTTCTTCACTCTCTGGCCATCAAGACATCATTCAATA
TCCCTACAACCATCAGTTAAAAGCATCTCCGGGCCATATGGTATCAGCAGTTCTGAATC
TCTGATCGATTACATGGCGTTTAAAGTCAAATAATGTTGTGAATCAACAAGGCTTTGAGTT
TCCTGAGGTGTCAAAGGAAATCAAGAAGGTGGTGAAGAAGGACCGACATAGCAAGATTCA
AACGGCACAAGGGATTAGAGACAGGAGGGTTAGGCTTTTATTGGGATTGCTCGCCAATT
CTTTGATCTTCAGGATATGTTGGGGTTTGATAAAGCTAGTAAACGTTAGACTGGCTGCT
CAAGAAGTCAAGAAAAGCCATCAAAGAGGTGCTACAAGCAAAAACCTCAACAATGATGA
TGAAGATTTTGGAAACATTGGAGGCGATGTAGAACAAGAAGAGGAGAAGGAGGAGGATGA
CAATGGCGATAAGAGCTTCGTGTATGGTTTGAGCCCCGGGTACGGTGAAGAAGAAGTGGT
ATGTGAGGCCACGAAGGCAGGATAAGAAAGAAGAAGAGTGAGTTGAGAAACATCTCATC
AAAGGGGCTAGGAGCCAAAGCTAGAGGAAAAGCAAGGAGCGAACAAAAGAGATGATGGC
CTATGATAATCCAGAGACTGCCTCTGATATTACACAATCTGAAATCATGGACCCATTCAA
GAGGTCTATAGTCTTCAATGAAGGAGAAGATATGACACACCTTTTCTACAAGGAACCAAT
CGAGGAGTTTGATAATCAAGAATCTATCTTAACCAATATGACTCTACCAACGAAGATGGG
TCAAAGTTACAATCAAAATAATGGGATACTTATGTTGGTAGATCAGAGTTCTAGCAGCAA
CTATAATACATTTCTGCCTCAAAATTTGGATTATAGTTATGATCAAAACCCTTTTTCATGA
CCAAACCTTATATGTAGTACCGACAAAAATTTCCCAAGGTTTCTTATAAATCTCGAC
AGTTTTGAAGGACTATGCATGATCAAGTTTAAACATGTAAGCCAATATAGTCCCTTATTC

CTCTGAATGTATACAAAATCTATAGTTATGTATATCTGTTCTTTTAAACGTATCTTTAT
TGATCTTCTGTGCCTTGATCAAAATGTCAATTTTAAAGATTCAGTTTGTGTAATATTTAG
CTACAACCTTTAAGTGGTATTATTGTAACCTTTTGAACATATATATTTGAAGATGAATAA
GAACATGTTTATATAAAAA

>G615 Amino Acid Sequence (domain in AA coordinates:88-147)
MSSSTNDYNDGNNGVYPLSLYLSSLGHQDIIHNPYNHQLKASPGHMSAVPESLIDYM
AFKSNVNVNQGFEPFEVSKEIKKVVKDRHSKIQTAAQGI RDRRLFIGIARQFFDLQD
MLGFDKASKTLDWLLKSKRKAIKEVVQAKNLNNDDEDFGNIGGDVEQEEKEEDDNGDKS
FVYGLSPGYGEEEVCEATKAGIRKKKSELNRNISSKGLGAKARGKAKERTKEMMAYDNPE
TASDITQSEIMDPFKRSIVFNEGEDMTHLFYKEPIEEFDNQESILTNMTLPKMGQSYNQ
NNGILMLVDQSSSNYNTFLPQNLDYSYDQNPFDQTLVYVTDKNFPKGFL*

>G732 (73..588)
AAAAAAACCAACATAAAACATAAAACTCTGTCCTTTTTTGTCTTCTTGTAACCTTTTCT
TGTTAAAAATCAATGGCGTCATCTAGCAGCACATACCGGAGCTCAAGCTCTTCCGACGGT
GGTAATAATAACCCGTCGGACTCCGTCGTCACCGTCGACGAACGAAAACGTAAAAGAATG
TTATCGAACAGAGAATCTGCACGTCAGGTCAAGGATCGGTAAACAGAAACACGTTGATGAT
CTAACGGCTCAGATCAATCAGCTATCAAACGACAACCGTCAGATCTTGAACAGCCTCACC
GTAACATCTCAGCTTTACATGAAGATCCAAGCCGAGAACTCTGTTCTCACCGCTCAGATG
GAGGAGCTTAGCACCAGACTCCAATCTCTCAACGAGATCGTTGATCTTGTTCATCAAC
GGTGCAGGATTTGGTGTGACCAAGATCGACGGCTGTGGTTTTGATGATCGTACGGTTGGG
ATCGACGGATATTACGATGATATGAATATGATGAGTAATGTTAATCATTGGGGTGGTTCG
GTTTACACTAACCAACCCATTATGGCTAATGATATCAATATGTATTGATTAATAAAATTA
ATTAATAAATTAGATGCCCTTTTGTCTTTTATTTTAAATTTAGCCCATTTTGGT
GTTTTTGGGTGGTGTGATGATGTAATTATAGTACATGCATCTTTGATTGGTTGGAAGGA
TAAATATAAACTTTATATATATATTGGGGCATATATATATGAGTTGTACTTTGCATGTAT
TGGTGTGTGTTTTGTTATAATTATATGATTATATATGTTTATGTTAAAAA

>G732 Amino Acid Sequence (domain in AA coordinates: 31-91)
MASSSTYRSSSSSDGNNNPSDSVTVDERKRRMLSNRESARRSRMRKQKHVDDLTAQ
INQLSNDNRQILNSLTVTSQLYMKIQAENSVLTAQMEELSTRQLSLNEIVDLVQSNAGF
GVDQIDGCGFDDRTVGIDGYDDMNMMSNVNHWGGSVYTNQPIMANDINMY*

>G988 (1..1338)
ATGCTTACTTCCTTCAAATCCTCTAGCTCCTCCTCCGAAGATGCCACCGCTACCACCACC
GAGAATCCTCCTCCTTTGTGCATCGCCTCCTCCTCGGCCGCAACCTCCGCTCACATCAC
CTCCGTCGTCTTCTTTTACCAGCTGCGAATTTTCGTCTCCAGTCAAACCTTACCAGCGCT
CAAACTTACTCTCAATCCTTCCCTTAACTCTTCTCCTCACGGCGACTCCACCGAGCGA
CTTGTACACCTCTTCACTAAAGCCTTGTCCGTACGAATCAACCGTCAGCAACAAGATCAG
ACGGCTGAAACGGTTGCCACGTGGACGACGAACGAAATGACGATGAGTAACTCCACGGTG
TTCACGAGCAGTGTATGCAAAGAACAGTTCCTTGTTCGAACCAAGAACAAATTCTGAC
TTCGAGTCTTGTACTATCTTTGGCTAAACCAACTAACGCCGTTTATTCGGTTCGGTCAT
TTAACGGCGAACCAAGCTATCCTCGACGCGACGGAGACAAACGATAACGGAGCTCTACAT
ATACTTGATTTAGATATATCACAAGGACTTCAATGGCCTCCATTGATGCAAGCCCTAGCA
GAGAGTCAATCAAACCTAGCAGTCCACCTCCATCTCTCCGCATAACCGGATGCGGTCGA
GATGTAACCGGATTAAACCGAATGGAGACCGGTTAACCGGTTGCTGACTCTTTAGGT
CTCCAATTCCAGTTTACACGCTAGTGATCGTAGAAGAAGATCTCGCCGGAATTTTGCTA
CAGATCCGATTGTTAGCTCTCTCAGCCGTACAAGGAGAGACCATTGCCGTCAATTGTGTT
CACTTCTCCACAAAATATTTAACGACGATGGAGATATGATCGGTCACTTCTTGTACGG
ATCAAGAGCTTAACTCTAGAATCGTTACAATGGCAGAGAGAGAACTAATCATGGAGAT
CACTCGTTCTTGAATAGATTCTCTGAGGCAGTGGATCATTACATGGCGATCTTTGATTCTG
TTGGAAGCGACGTTGCCGCCAAATAGCCGAGAGAGACTAACCTTAGAGCAACGGTGGTTC
GGTAAGGAGATTTGGATGTTGTGGCGCGGGAAGAGACGGAGAGAGAAAGCAAAGACATCGG
AGGTTTGAGATTTGGGAAGAGATGAGAGGTTTGGTTTACGTTAACGTTCTTATTGGA
AGCTTTGCTTTGTCTCAAGCTAAGCTTCTTCTTAGACTTCATTATCCTTCAGAAGGTTAT
AATCTTCAGTTTCTTAAACAATCTTTGTTTCTTGGCTGGCAAAATCGTCCCCTCTTCTCC
GTTTCGTCTGGAAATGA

>G988 Amino Acid Sequence (domain in AA coordinates:178-195)
MLTSFKSSSSSEDATATTTENPPPLCIASSSAATSASHHLRRLFTAAANFVSQSNFTAA
QNLLSLSLNSPPHGDSTERLVHLFTKALSVRINRQQDQTAETVATWTTNEMTMSNSTV

FTSSVCKEQFLFRTKNNNSDFESCYLWLNQLTPFIRFGHLTANQAILDATETNDNGALH
ILDLDISQGLQWPPLMQALAERSSNPSSPPPSLRITGCGRDVTGLNRTGDRLTRFADSLG
LQFQFHTLVIVEEDLAGLLQIRLLALSAVQGETIAVNCVHFLHKIFNDDGDMIGHFLSA
IKSLNSRIVTMAEREANHGDSFLNRFSEAVDHYMAIFDSLEATLPPNSRERLTLEQRWF
GKEILDVVAEETERKQRHRRFBIWEEMMKRFGFVNVPISFALSQAKLLRLHYPSEGY
NLQFLNNSLFLGWQNRPLFSVSSWK*

>G1519 (1..1146)

ATGAGGCTTAATGGGGATTCGGGTCCGGGTCAGGATGAACCCGGTTCGAGCGGGTTTCAC
GGCGGAATCAGACGATTCCCGTTAGCAGCTCAGCCGGAGATTATGAGAGCTGCTGAGAAA
GACGATCAATACGCTTCTTTTCATCCACGAAGCTTGCCGCGATGCCTTCCGACACCTTTTC
GGTACAAGAATCGCTCTTGCTTACCAGAAGGAGATGAAGCTACTTGGACAGATGCTTTAC
TATGTTCTTACGACAGGTTACAGGCAACAACTTTAGGAGAGGAATATTGTGACATTATA
CAGGTTGACAGGGCTTATGGACTCTCTCTACACCAGCTAGACGTGCTTTGTTTATATTG
TACCAGACCGCAGTTCCATATATCGCAGAGAGAATTAGCACTCGAGCTGCTACGCAAGCA
GTCACCTTTGATGAGTCTGATGAGTTTTTTGGTGATAGTCATATCCACTCACCAAGAATG
ATAGATCTTCCATCTTCATCTCAAGTTGAACTTCAACTTCTGTAGTATCTAGGTTAAAC
GATGACTTATGAGATCGTGGCACCGAGCTATTACGCGATGGCCTGTGGTCTTCTCTGTT
GCCCAGCAAGTCTTACAACGTGTTTTCGCTGCCAATCTGATGCTCTTCTACTTTGAAGGT
TTTTTATTATCATATATCGAAACGTGCATCCGGGGTTCGTTATGTTTTTCATAGGAAAGCAA
CTGAATCAGAGACCTAGATACCAAATTCTTGGGGTTTTCTTCTAATCCAATTGTGCATC
CTTGCTGCTGAGGGCTTGCGTCGGAGTAATTTGTCTATCTACTAGCTCCATTACAGCAG
GCTTCTATAGGATCTTATCAAACTTCAGGAGGGAGAGGTTTACCTGTTTTAAATGAAGAG
GGGAATTTGATAACTTCGGAAGCTGAAAAGGGAACTGGTCTACCTCCGATTCAACTTCA
ACGGAGGCAGTAGGGAAATGCACTCTCTGCTTAAGCACCCGTCAGCACCCAACGGCCACT
CCTTGTGGTCATGTGTTTTGTTGGAGCTGCATTATGGAATGGTGCAACGAGAAGCAAGAA
TGCCCTCTTTGTGCAACGCCCAATACCCATTCAAGTTTGGTTGTTTGTATCATTTCTGAT
TTTTAG

>G1519 Amino Acid Sequence (domain in AA coordinates: 327-364)

MRLNGDSGPGQDEPGSSGFHGGIRRFPLAAQPEIMRAAEKDDQYASFIHEACRDAFRHLF
GTRIALAYQKEMKLLGQMLYYVLTGSGQQTLGEEYCDIIQVAGPYGLSPTPARRALFIL
YQTAVPYIAERISTRATQAVTFDESDEFFGDSHIHSPRMIDLPSSSQVETSTSVVSRNLN
DRLMRSWHRAIQRPVVLPVAREVLQLVLRANLMLFYFEGFYHISKRASGVRYVFIGKQ
LNQRPYQILGVFLLIQLCILAAEGLRRSNLSSITSSIQQASIGSYQTSGGRGLPVLNEE
GNLITSEAEKGNWSTSDSTSTEAVGKCTLCLSTRQHPTATPCGHVFCWSCIMEWCNEKQE
CPLCRTPNTHSSSLVCLYHSDF*

>G374 (1..1359)

ATGGACAACAAAATGATCAGGATATTGATGTTAGATCAGTGGTTGAAGCTGTTTCCGCC
GATCTTTCTTTGGTGCTCCCTCTATGTGGTTGAGAGCATGTGCATGCGCTGCCAAGAA
AATGGAACAACAGATTTCTATTGACCTTAATTCCTCACTTCAGAAAGGTTCTAATATCT
GCATTTGAATGTCCGCATTGCGGGGAAAGGAATAATGAAGTTCAGTTCGCAGGCGAGATT
CAACCCCGTGATGCTGTTACAATCTAGAGGTTCTAGCTGGTGATGTGAAGATATTTGAC
CGGCAAGTTGTGAAATCTGAATCAGCCACTATTAAGATTCCTGAACCTGGATTTTGAGATT
CCACCAGAGGCCCCAACGTGGAAGTTTGTCTACTGTGGAAGGGATATTAGCACGGGCTGCT
GATGAAC TGAGTGCCCTTCAAGAAGAACGCAAGAAAGTTGATCCTAAAAC TGCTGAAGCA
ATAGACCAATTCTTGTCAAAAC TGAGAGCTTGCTGCTAAAGCAGAGACATCCTTCACCTTC
ATTTTGGATGATCCTGCTGGAAACAGTTTCATTGAGAACCACATGCTCCATCACCAGAT
CCCTCTCTAACCATCAAATCTATGAGCGAACACCAGAGCAACAAGCAACACTTGGATAT
GTTGCTAACCCATCTCAGGCTGGACAATCAGAAGGAAGCCTTGGCGCACCTGTGATGACT
TTCCCTTCAACTTGGCGAGCATGTACGGAGCCGTGTGAGACACGGATGTTCAAAATAGAA
ATCCCGTACTTTACGGAAGTTATTGTGTCATGGCATCTACATGTGACAGTTGTGGCTATCGT
AATTCTGAGTTGGAAGCCTGGTGGTGCAATTCCTGAAAAGGGAAAGAAGATTACTCTCTCT
GTGAGGAACATTACAGACCTTAGCCGAGATGTTATCAAGTCGGACACTGCAGGAGTGATA
ATCCCAGAACTTGATCTGGAGCTAGCTGGTGGTACACTGGTGGAATGGTAACAACAGTT
GAAGGGTTGGTTACACAGATCAGAGAAAGCCTAGCGAGAGTTCACGGATTCACTTTTGGT
GATAGTATGGAAGAGAGTAAGTTGAACAAATGGAGAGAATTTGGAGCCAGGCTCACTAAG
CTCCTAAGCTTTGAACAGCCGTGGACATTGATTCTTGATGATGAATTAGCAAATTCCTTT
ATTGCACCAGTAACAGATGATATCAAAGATGACCATCAGCTCACATTTGAAGAGTACGAG

AGGTCATGGGATCAAAACGAGGAGTTGGGTCTCAACGACATAGATACTTCTTCAGCTGAT
 GCTGCTTATGAATCCACAGAGACGACTAAATTACCTTAA
 >G374 Amino Acid Sequence (domain in aa coordinates: 35-67, 245-277)
 MDNKNQDIDVRSVVEAVSADLSFGAPLYVVESMCMRCQENGTRFLLTLIPHFRKVLIS
 AFECPHCGERNNEVQFAGEIQPRGCCYNLEVLAGDVKIFDRQVVKSESATIKIPELDFEI
 PPEAQRGSLSTVEGILARADELALQEERKKVDPKTAEAIDQFLSKLRACAKAETSFTF
 ILDDPAGNSFIENPHAPSPDPSLTIKFYERTPEQQATLGIVANPSQAGQSEGLGAPVMT
 FPSTCGACTEPCETRMFKIEIPYFQEVIVMASTCDSCGYRNSCLKPGGAIPKGGKITLS
 VRNITDLSRDVIKSDTAGVIEPELDLELAGGTLGGMVTTVEGLVTQIRESLARVHGFTEG
 DSMEESKLNKWRFEFARLTKLLSFEQPWTLILDDELANSFIAPVTDDIKDDHQLTFEEYE
 RSWDQNEELGLNDIDTSSADAAYESTETTKLP*
 >G877 (397..2460)
 CAAAGATTAGACTAATCCGACTGTGTTTTTAAATCAATCATCATTTTATTAGGGGAGAGA
 AGTTGTAAAGTTTGGATTTTTTTTCTGGGTTTTTCTGTGAGACCCAGAAGAAGAACAG
 AGAGAGGAAGAAGGAGAAGAAAAAATATCTCTTCTCTCCGGCTTTCAACAAAATCTCT
 CTTTTTTCCTTCATCAGTGTTAAATTCCGATCCGGGTCGGGTGGGTTTTCGGTTTTTGGT
 GTTCGGATCAGAGCACAGTTGGATGTTAGCGACGGAAGTGGAGTTTCAGTTTGCGGCTG
 CGGCGGCTGTGACGGTGTGTGTGTCTCTTTTATCAATCAGGAGTTTCATCACAG
 TTTGATCAGAGATTAGCCAAATCTTGGATACTAAATGGCTGGTTTTGATGAAAATGTT
 GCTGTGATGGGAGAAATGGGTGCCTCGTAGTCTTAGTCCCGGACACTTTTCTCCTCTGCT
 ATTGGAGAAGAGAAGAGCTCGAAACGTGTTCTTGAAGAGAGTTATCTTTGAATCATGGT
 CAAGTTATTGGTTTAGAAGAAGACACTAGTAGTAATCATAACAAGGATCTTCACAAAGC
 AATGTTTTTTCGAGGTGGTCTCAGTGAAAGAATTGCTGCAAGAGCTGGATTTAATGCTCCA
 AGGTTGAACACTGAGAATATCCGCACCAACACCGACTTTTCCATTGACTCTAACCTTCGA
 TCTCCTTGCTTAACCATCTCTTCTCCTGGCCTTAGCCCTGCAACACTCTTGAATCTCCT
 GTTTTCCTTTCTAACCCATTGGCTCAACCTTCTCCAACCTACCGGGAATTTCCATTTCTT
 CCTGGTGTAAATGGTAATGCATTGTCTTCTGAGAAAGCGAAAGACGAGTTCTTTGATGAT
 ATTGGAGCATCATTAGCTTCCATCCTGTTTCAAGATCATCTTCTCTTTCTTCCAAGGC
 ACAACAGAGATGATCAGTTGATTATGGTAACACAAATAGATCTTCTTCTCATCAA
 TCCGCAGAAGAAGTAAACCTGGCTCTGAAAACATAGAAAGCTCCAATCTTTATGGGATT
 GAAACTGACAATCAAAACGGGCAGAACAAAGACATCTGATGTCACTACAAACACCGTCTT
 GAAACCGTGGATCATCAAGAGGAAGAAGAAGAGCAAAGACGCGGTGATTCGATGGCTGGT
 GGTGCGCCTGCAGAGGATGGATATAACTGGAGGAAATACGGACAAAAGTTGGTCAAAGGA
 AGTGAGTATCCGCGAAGCTATTACAAGTGCACAAACCCGAATTGTCAGGTGAAGAAGAAA
 GTTGAGAGATCAAGGGAAGGTCACATCACAGAGATTATATACAAAGGAGCTCATAATCAT
 CTTAAACCTCCACCTAATCGCCGCTCAGGGATGCAAGTAGATGGAACGAAACAAGTTGAA
 CAACAACAACAACAGAGAGATTCTGCTGCAACGTGGGTTAGTTGTAATAACACTCAACAA
 CAAGGTGGAAGCAATGAGAACAATGTGCAAGAGGGATCTACGAGATTGAGTATGGAAC
 CAATCTGGATCAATTCAAGCTCAAACCGGAGGTCAATACGAGTCAAGGTGATCCTGTGGTT
 GTGGTTGATGCTTCTTCAACATTCTTAATGATGAAGATGAAGATGATCGAGGGACACAT
 GGAAGTGTCTTTGGGTTACGATGGAGGAGGAGGAGTGGGGGAGGAGAAGGAGATGAA
 TCAGAGTCGAAAAGAAGGAACTAGAAGCTTTTGCAGCAGAGATGAGTGGATCAACAAGA
 GCCATACGTGAGCCAAGAGTTGTTGTGCAGACAACGAGTGATGTTGACATTCTTGATGAT
 GGTATATCGCTGGCGAAAATATGGTCAGAAAAGTTGTCAAAGGCAATCCAAATCCAAGGAGT
 TATTACAAATGCACAGCTCCAGGATGTACAGTGAGGAAACATGTTGAAAGAGCTTCTCAT
 GATCTCAAATCCGTTTATAACAACCTACGAAGGCAAACATAACCATGACGTCCCGCTGCA
 CGCAACAGCAGCCACGGAGGCGGTGGTATAGTGGTAACGGTAACAGCGGCGGTTTCAGCC
 GCAGTTTCTACCATTAACACAACGGTCATCACTCAGAGCCGCCACGTGGGAGATTTCGAC
 AGACAAGTCACAACATAACAATCAGTCTCCTTTTAGCCGTCCCTTTAGCTTTTCAGCCACAT
 TTGGGTCCTCCTTCTGTTTCTCCTTCGGTTTTAGGACAAACCGGTTTTGGTTAATCTTTCA
 ATGCCTGGTTTTAGCGTATGGTCAAGGGAATGCGGGTTTGCCTCACCCGTATATGACA
 CAACCGGTTGGGATGAGTGAAGCAATGATGCAGAGAGGGATGGAACCAAGGTTGAACCG
 GTTTTCAGATTGAGGACAATCGGTATATAACCAGATCATGAGTAGATTACCTCAGATTGGA
 AATTTACTCTTCTTCTTCTTCTGCAATTTGGTCACTCCTTATAATAACTTTTAATTTT
 TGCTTCTTCTTCTTCTTCTTATTTGTTTCAAACCTTTGGGGAAGGTAAGGCTGTTTT
 ATTGTTAAAAAATAAAAAAAAAA
 >G877 Amino Acid Sequence (domain in AA coordinates: 272-328, 487-603)

MAGFDENVAVMGEWVPRSPSPGTLFSSAIGEEKSSKRVLERELSLNHGQVIGLEEDTSSN
 HNKDSSQSNVFRGGLSERIAARAGFNAPRLNTEINRTNTDFSIDSNLRSPCLTISSPGLS
 PATLLESPVFLSNPLAQPSPTTGKFPFLPGVNGNALSSSEKAKDEFFDDIGASFSFHPVSR
 SSSSFFQGTTEMMSVDYGNYNRRSSSHQSAEEVKPGSENISSNLYGIETDNQNGQNKTS
 DVTNTSLETVDHQEEEEQRRGDSMAGGAPAEADGYNWRKYGQKLVKGSEYPRSYKCTN
 PNCQVKKKVERSREGHITEIIYKGAHNHLKPPNRRSGMQVDGTEQVEQQQQQRDSAATW
 VSCNNTQQQGGSNENNVEEGSTRFEYGNQSGS IQAQTGGQYESGDPVVVDASSTFSNDE
 DEDDRGTHGSVSLGYDGGGGGGGGEGDESESKRRKLEAFAAEMSGSTRAIREPRVVQTT
 SDVDILDDGYRWRKYGQKVVKGNPNPRSYKCTAPGCTVRKHVERASHDLKSVITTYEGK
 HNHDVPAARNSSHHGGGDSGNGNSGGSAAVSHHYHNGHHSEPPRGRFDRQVTTNNQSPFS
 RPFQPHLGPSPGFSFGLGQTLGNLSMPGLAYGQKGMPGLPHPYMTQPVGMSEAMMQR
 GMEPKVEPVSDSGQSVYNQIMSRLPQI*

>G1000 (1..954)

ATGGGAAGACCTCCTTGTGTGACAAGTCCAATGTCAAGAAAGGTCTCTGGACCGAGGAA
 GAAGACGCTAAGATCCTTGTCTATGTTGCTATCCATGGTGTAGGAACTGGAGCTTGATC
 CCCAAAAAGCAGGTCTGAATCGATGTGGAAGAGCTGTAGACTAAGATGGACTAATTAC
 TTAAGACCTGACCTTAAACATGACAGCTTCTCTACCCAAGAAGAAGAGCTTATCATTGAG
 TGTATAGAGCCATTGGCAGCAGGTGGTCTTCCATTGCACGAAAGCTTCCAGGAAGAACG
 GATAATGATGTGAAGAATCACTGGAACACAAAGCTGAAGAAGAAGCTGATGAAAATGGGG
 ATAGACCCGGTGACTCATAAACCGGTTTCTCAACTCCTTGCAAGATTGAGAAACATTAGC
 GGCCATGGAATGCATCCTTCAAAACAGAACCATCTAACAACCTCTATACTCACACAATCC
 AACTCAGCTTGGGAAATGATGAGAAACACAACAACAAACCATGAGAGTTATTACACCAAC
 TCTCCAATGATGTTTACAAATTCCTCTGAGTACCAACTACTCCATTTCTATAGC
 CATCCAAATCATCTGCTCAATGGAACCACATCTTCATGCTCTTCCTCATCATCTTCTACT
 AGTATCACTCAGCCAAACCAAGTACCTCAAACACCGGTTACTAATTCTACTGGAGCGAT
 TTCCTTCTCTCGGACCCGGTTCTCAAGTAGTGGGATCCTCAGCTACTAGCGACCTCACT
 TTTACGCAGAACGAACATCATTTCAACATCGAAGCCGAATACATCTCTCAAAACATCGAT
 TCAAAGGCCCTCGGGAACATGTCAATCCGCGAGTTCTTCGTTGACGAAATACTAGATAAA
 GACCAAGAGATGTTGTACAGTTTCTCAACTCTTGAATGATTTGATTATTAG

>G1000 Amino Acid Sequence (domain in AA coordinates: 14-117)

MGRPPCCDKSNVKKGLWTEEDAKILAYVAIHGVGNWSLIPKAGLNRCKGKSCRLRWNTY
 LRPDLKHDSTFQEEELIIECHRAIGSRWSSIARKLPGRTDNDVKNHWNKLLKKLMMKG
 IDPVTHKPVSQLAEFRNISGHGNASFKTEPSNNSILTQNSAWEMMRNTTNNHESYYTN
 SPMFTNSSEYQTFHFYSHPNHLLNGTTSSCSSSSSSTSITQPNQVPQTPVTNFWSD
 FLLSDPVPQVVGSSATSDLTFTQNEHHFNIEAEYISQNIIDSKASGTCHSASSFVDEILDK
 DQEMLSQFPQLNDFDY*

>G1067 (436..1371)

TCTCAAGCTTCTCTCTCTCTTTTTCCTATAGCACATCAGAATCGCTAAATACGACTCCT
 ATGCAAAGAAGAGCTACTTCTTCTCTTGGCCCTAATTAATCTACCTAACTAGGGTTTCC
 TCTTACCTTTTCTAGAGAGAGATCATTTAACAATAAGTCACCTTTTTTATATCTTTTGCTTC
 GTCTTTAATTTAGTTCTGTTCTTGGTCTGTTTCTATATTTGTCTGGCTTGCGTAACCGAT
 CACACCTTAATGCTTTAGCTATTGTTTCTCAAATCATGAGTTTGGACTTCTCGATCTG
 AGTTTTCTTTTCTCTCTTTACGCTCTTCTTACCTAGCTACCAATATATGAACGAGCAG
 GATCAAGAATCGAGAAATTGATTGAGCTGGCGAATAAGCAGTGGTGGGATAGGGAATTA
 GTAGATGCGGCGGCGATGGAAGGCGGTTACGAGCAAGGCGGTGGAGCTTCTAGATACTTC
 CATAACCTCTTTAGACCGGAGATTACCAACCAACAGCTTCAACCGCAGGGCGGGATCAAT
 CTTATCGACCAGCATCATCATCAGCACCAGCAACATCAACAACAACAACCGTCCGGAT
 GATTCAAGAGAACTGACCATTCAAACAAAGATCATCATCAACAGGGTCGACCCGATTCA
 GACCCGAATACATCAAGCTCAGCACCAGGAAAACGTCCACGTGGAGCTCCACCAGGATCT
 AAGAACAAAGCCAAAGCCAGCATATAGTAACCTCGTGATAGCCCCAACGCGCTTAGATCT
 CACGTTCTTGAAGTATCTCTGGAGCTGACATAGTTGAGAGTGTTCACGTACGCTAGG
 AGGAGAGGGAGAGCGCTCTCCGTTTGTAGGAGGAAACGGCACCGTATCTAACGTCACTCTC
 CGTCAGCCAGTCACTCCTGGAATGGCGGTGGTGTGTCCGAGGAGGAGGAGTTGTGACT
 TTACATGGAAGGTTGAGATTCTTTCGCTAACGGGAGTGTTCGACCTCTGACCCG
 CCTGGTGCCGGTGGTTGTCTATATTTTAGCCGGAGGGCAAGGTGAGGTGGTGGGAGGA
 AGCGTTGTGGCTCCCTTATTGCATCAGCTCCGTTTATACTAATGGCGGCTTCGTTCTCA
 AATGCGGTTTTTCGAGAGACTACCGATTGAGGAGGAGGAAGAAGAAGGTGGTGGTGGCGGA

GGAGGAGGAGGAGGAGGGCCACCGCAGATGCAACAAGCTCCATCAGCATCTCCGCCGTCT
GGAGTGACCGGTGAGGACAGTTAGGAGGTAATGTGGGTGGTTATGGGTTTTCTGGTGAT
CCTCATTTGCTTGGATGGGAGCTGGAACACCTTCAAGACCACCTTTTTAATTGAATTTT
AATGTCCGAAATTTATGTGTTTTATCATCTTGAGGAGTCGCTTTCTTTGGGATATT
TGGTGTTAATGTTTAGTTGATATGCATATTTT

>G1067 Amino Acid Sequence (domain in AA coordinates: 86-93)
MEGGYEQGGGASRYFHNLFPRPIHHQQLQPQGGINLIDQHHHQHQQHQQQQPSDDSRES
DHSNKDHHQQGRPDSDPNTSSSAPGKRPRGRPPGSKNKAKPPIIVTRDSPNALRSHVLEV
SPGADIVESVSTYARRRGRGVSVLGGNGTVSNVTLRQPVTGNGGGVSGGGGVVTLHGRF
EILSLTGTVLPPPAPPAGGLSIFLAGGQGVVGGSVVAPLIASAPVILMAASFNAVFE
RLPIEEEEEGGGGGGGGGPPQMQAAPSASPPSGVTGQQQLGQNVGGYGFSGDPHLLG
WGAGTPSRPPF*

>G1075 (19..876)
TTTGTGTTTGGTGTGGCATGGCTGGTCTCGATCTAGGCACAACCTTCTCGCTACGTCCAC
AACGTCGATGGTGGCGGCGCGGACAGTTCACCACCGACAACCACCACGAAGATGACGGT
GGCGCTGGAGGAAACCACCATCATCACCATCATAATCATAATCACCATCAAGGTTTAGAT
TTAATAGCTTCTAATGATAACTCTGGACTAGGCGGCGGTGGAGGAGGAGGGAGCGGTGAC
CTCGTCATGCGTCGGCCACGTGGCCGCTCAGCTGGATCGAAGAACAACCGAAGCCGCCG
GTGATTGTACGCGCGAGAGCGCAACACTCTTAGGGCTCACATTCTGAAGTTGGAAGT
GGCTGCGACGTTTTCGAATGTATCTCCACTTACGCTCGTCGGAGACAGCGCGGATTTGC
GTTTTATCCGGGACGGGAACCGTCACTAACGTCAGCATCCGTCAGCCTACGGCGGCCGGA
GCTGTTGTGACTCTGCGGGGTACTTTTGAGATTCTTCCCTCTCCGGATCTTTCTTCCG
CCACCTGCTCCTCCAGGGCGACTAGCTTGACGATATTCCTCGCTGGAGCTCAAGGACAG
GTCGTCGGAGGTAACGTAGTTGGTGAGTTAATGGCGCGGGGCCGGTAATGGTCATGGCA
GCGTCTTTTACAAACGTGGCTTACGAAAGGTTGCCTTTGGACGAGCATGAGGAGCACTTG
CAAAGTGGCGGCGCGGAGGTGGAGGGAATATGTACTCGGAAGCCACTGGCGGTGGCGGA
GGGTTGCCTTTCTTTAATTTGCCGATGAGTATGCCTCAGATTGGAGTTGAAAGTTGGCAG
GGGAATCACGCCGCGCGCGGTAGGGCTCCGTTTTAGCAATTTAAGAACTTTAATTGTTT
TTTCCACTTTTTTGTGTTTTCTCCGAATTTTATGAAATTATGATTTAAGAAAAAACGAT
ATTGTTTCATGATTGACCTCTTACTGCATGGTTTCTTCTATTGGGTTAATTGGCTAGCT
CATAAGAATTGTTAATTTGGTTATTGTCATCAAATTTGCCACATATAAAGCTTCTAGC
AAAT

>G1075 Amino Acid Sequence (domain in AA coordinates: 78-85)
MAGLDLGTTSRYVHNVDGGGGGQFTTDNHHEDDGGAGGNHHHHHHNHNHHQGLDLIASND
NSGLGGGGGGSGDLVMRRPRGRPAGSKNKPVPVIVTRESANTLRAHILEVSGCDVFE
CISTYARRRQRGICVLSGTGTINVSIRQPTAAGAVVTLRGTFFELSLSGSFLPPPAPP
ATSLTIFLAGAQGVVGVGNVVELMAAGPVMVMAASFNTVAYERLPLDEHEEHLQSGGGG
GGGNMYSEATGGGGGLPFFNLPMSPQIGVESWQGNHAGRAPF*

>G1266 (62..718)
CAATCCACTAACGATCCCTAACCGAAAAACAGAGTAGTCAAGAAACAGAGTATTTTTCTA
CATGGATCCATTTTAATTCAGTCCCCATTCTCCGGCTTCTCACCAGGAATATTCTATCGG
ATCTTCTCCAGATTCTTTCTCATCCTCTTCTTCTAACAATTACTCTCTTCCCTTCAACGA
GAACGACTCAGAGGAAATGTTTTCTACGGTCTAATCGAGCAGTCCACGCAACAAACCTA
TATTGACTCGGATAGTCAAGACCTTCCGATCAAATCCGTAAGCTCAAGAAAGTCAGAGAA
GTCTTACAGAGGCGTAAGACGACGGCCATGGGGGAAATTCGCGGCGGAGATAAGAGATT
GACTAGAAACGGTATTAGGGTTTGGCTCGGGACGTTGAAAGCGCGGAAGAGCGGCTTT
AGCCTACGATCAAGCTGCTTTCTCGATGAGAGGGTCTCGGCGATTCTCAATTTTTCGGC
GGAGAGAGTTCAAGAGTCGCTTTTCGAGATTAAATATACCTACGAGGATGGTTGTCTCC
GGTTGTGGCGTTGAAGAGGAAACACTCGATGAGACGGAGAATGACCAATAAGAAGACGAA
AGATAGTGACTTTGATCACCGCTCCGTGAAGTTAGATAATGTAGTTGTCTTTGAGGATTT
GGGAGAACAGTACCTTGAGGAGCTTTGGGGTCTTCTGAAAATAGTGGGACTTGGTGAAA
GATTAGGATTTGTATTAGGGACCTTAAGTTTGAAGTGGTTGATTAATTTTAAACCTAATA
TGTTTTTTGTTTGTCTAAATATTGATTCTATTGAGAAACATCGAAAACAGTTGTATGT
ACTTTTGTGATACTTGGCG

>G1266 Amino Acid Sequence (domain in AA coordinates: 79-147)
MDPFLIQSPFGFSPEYSIGSSPDSFSSSSNNYSLPFNENDSEEMFLYGLIEQSTQOTY
IDSDSQDLPIKSVSSRKSEKSYRGVRRRPWGKFAAEIRDSTRNGIRVWLGTFFESAEEAAL

AYDQAAFSMRGSSAILNFSERVQESLSEIKYTYEDGCSPPVVALKRKHSRRRMTNKKTK
DSDFDHRSVKLDNVVVFEDLGEQYLEELLGSSSENSGTW*

>G1311 (41..757)

AAGTATAATAACACAAAGAACAGAGTAAAAGAAAGAAAAATGGATTTTAAAGAGGAAGA
AACACTTCGTAGAGGGCCATGGCTCGAAGAAGAAGACGAACGGCTAGTGAAGGTCATTAG
TCTTTTGGGAGAACGTCGTTGGGATTCTTTAGCAATAGTTTCCGGTTTGAAGAGGAGTGG
TAAGAGTTGCGAGGCTAAGGTGGATGAACATCTGAATCCGACTCTGAAGCGTGGACCGAT
GAGTCAAGAAGAAGAGAGAATCATCTTTCAGCTCCATGCTCTATGGGGTAACAAGTGGTC
GAAGATTGCGAGAAGATTACCCGGTAGGACTGATAACGAGATAAAGAACTATTGGAGAAC
TCATTATAGAAAGAAACAGGAAGCTCAAACTATGGAAGACTCTTTGAGTGGAGAGGAAA
TACAGGAGAAGAATTGTTGCACAAGTATAAGGAAACAGAGATCACTAGGACAAAGACGAC
GTCTCAAGAACATGGTTTGTGTTGAAGTTGTGAGCATGGAAGTGGTAAAGAAGCCAAACGG
TGGTGTGGTGGTGAAGAGAAAGCTTCGGTGTATGAAATCACCGTATGAAAATCGGATTTTC
GGATTGGATATCAGAGATTTCTACTGACCAGAGTGAAGCAAATCTTTCAGAAGATCACAG
CAGCAATAGCTGCAGTGAGAACAAATATTAACTATTGGTACTTGGTGGTTTCAAGAGACTAG
GGACTTTGAGGAGTTTTCATGTTCTCTATGGTCATAATTCTAAAGTTGGTTTATTTACTT
TTTAAAAA

>G1311 Amino Acid Sequence (domain in AA coordinates: 11-112)

MDFKKEETLRGPPWLEEDERLVKVISLLGERRWDLSLAIVSGLKRSKSCRLRWNNYLN
TLKRGPMSEEBRIIFQLHALWGNKWSKIARRLPGRDNEIKNYWRTHYRKKQEAQNYGK
LFEWRGNTGEBLLHKYKETEITRTKTTTQEHGFVEVVSMEGKEANGGVGGRESFGVMKS
PYENRISDWISEISTDQSEANLSEHSSNSCSENNINIGTWWFQETRDFFESCSLWS*

>G1321 (72..803)

GTTCTTGATTGGTTTGGATCGGTATACCTTAGTTGATTACGTAATTAAATAGATCGGCGT
GAAGAAGAAAAATGATCATGTGCAGCCGAGGCCATTGGAGACCAGCTGAAGACGAGAAGC
TCAAGGATCTTGTGCAACAATACGGTCCTCACAATTGGAACGCCATTGCTCTCAAGCTTC
CTGGTCGCTCTGGTAAGAGTTGTAGATTGAGATGGTTTAAATCAATTGGATCCAAGGATCA
ACCGAAACCCCTTTCACGGAAGAAGAAGAAGAAAGACTTTTAGCGGCTCATCGGATCCATG
GGAACAGATGGTCCATCATCGCAAGGCTTTCCCTGGAAGAACTGATAACGCCGTCAAGA
ACCATTTGGCACGTCATCATGGCTCGTCGCACACGCCAAACCTCTAAGCCTCGTCTTCTTC
CCTCGACGACTTCGTCTTCTTCTTTAATGGCGAGTGAACAAATCATGATGAGTTCTGGTG
GTTATAATCATAATTATAGTTCCGATGATCGGAAGAAAATATTTCCAGCAGACTTTATAA
ATTTCCCTTACAAATCTCTCATATCAATCATCTTCACTTCTTAAAGGAGTTTTTCCCGG
GAAAGATCGCTTTAAGTCACAAAGCAAATCAGAGTAAGAAGCCTATGGAGTTCTACAATT
TTCTACAAGTAAACACAGATTCAAACAAGAGCGAGATTATAGATCAAGATTCAGGTCAAA
GCAAACGCAGTGATCGGACACCAACATGAAAGTCATGTTCCATTCTTCGACTTTTAT
CCGTTGGAAACTCTGCCTCTCTAGGATTAGTTTTTTTGCAGTAACCTCTAAATTTCTAGAT
TAACTATTTAGTCCGTATACGTACGAGATTATCTAGGTCGTTAGCATGTATGCTTGATGT
GTATAATCACTAACTAGTGAGCTATTACCTGCGAAAATTGTAAGAAAATACATAATGTT
GATGTATCACACATTCTCAATGTCTGTAAATTTCCATCGAGTTGTTAACTATCAAAGTT
ATCCGTTTGAAAAA

>G1321 Amino Acid Sequence (domain in AA coordinates: 4-106)

MIMCSRGHWRPAEDEKLKDLVEQYGPNNWNAIALKLPRSGKSCRLRWNNQLDPRINRNP
FTEEEERLLAAHRIHGNRWSIIARLFPGRDNAVKNHWHVIMARRTRQTSKPRLLPSTT
SSSSLMASEQIMSSGGYNNHYSDDDRKKIFPADFINFPYKFSHINHLHFLKEFFPGKIA
LSHKANQSKKPMEFYNFLQVNTDSNKSEIIDQDSGQSKRSDSTKHESHVPFFDLSVGN
SAS*

>G1326 (32..784)

CGACGGTACGGTGGAGATAGAGATAGCATCCATGGAGATGTCTAGAGGAAGCAACAGTTT
TGACAATAAGAAGCCTAGTTGCCAAAGAGGTCACTGGAGACCTGTTGAAGATGACAATCT
CCGGCAACTCGTTGAACAATACGGTCCCAAGAACTGGAATTTTATGCTCAACATCTCTA
TGGAAGATCAGGGAAGAGCTGTAGATTAAAGATGGTACAACCAACTTGATCCAAACATCAC
CAAGAAAACCTTTCACCGAGGAGGAAGAAGAGAGACTGCTTAAAGCTCATCGGATCCAAGG
GAATCGTTGGGCCTCATAGCCCGACTGTTCCCGGGAGGACCGACAACGCTGTCAAAAA
CCATTTTCATGTCATCATGGCTAGACGCAACGGGAAAACCTTCTCTCCACAGCTACTTC
TACGTTCAACCAAACTTGGCATACTGTTTGGAGCCCTAGTTCTAGTCTTACAAGGCTAA
TAGATCCCATTTCCGGGCTATGGAGGTATCGAAAGGATAAGAGTTGCGGTCTCTGGCCTTA

CTCTTTTGTTCACCACTACGAATGGTCAATTTGGATCTTCATCTGTCTCTAACGTACA
CCACGAAATTTATCTTGAGAGGAGAAAGTCGAAAGAGTTGGTGGATCCTCAGAATTACAC
ATTTTCATGCAGCCACACCAGATCATAAGATGACTTCAAATGAAGATGGACCATCCATGGG
AGATGATGGTGAGAAGAAGCATGTTACTTTTCATTGATTTTCTTGGTGTGGATTAGCTTC
TTAGGTTATAACATCACAAGTCAAAGCTTTTAAGGGTTTCTATCATTAGGGTTAGGCATC
ATTTTCAGCCTTTTGCTTCCTTAAACTCTCATATGGATCT

>G1326 Amino Acid Sequence (domain in AA coordinates: 18-121)
MEMSRGSNSFDNKKPSCQRGHWRPVEDDNLRLQVEQYGPKNWNFIAQHLYGRSGKSCRLR
WYNQLDPNITKKPFTEEEERLLKAHRIQGNRWASIALRFPGRTDNAVKNHFMVIMARRK
RENFSSTATSTFNQTWHTVLSPPSSSLTRLNRSHFGLWRYRKDKSCGLWPYSFVSPPTNGQ
FGSSSVSNVHHEIYLERRKSKELVDPQNYTFHAATPDHKMTSNEGSPMGDDGEKNDVTF
IDFLGVGLAS*

>G1367 (128..1567)

TCCTTCCACAAAACCTTTTTTAATTTTATCTGAAAAATTAAACAACCGAAACAAAAA
AAAACATAAAATCAAAAATCTCATCACCTTCCTTGCTCTGTATTTTCTCTCTCACTAA
ATCCTCCATGGATCCTTCTCTCTCTGCAACCAATGATCCTCATCATCCTCCTCCTCA
GTTTCACATCTTTCCTCCTTTCACCAACACCAACCCCTTCGCCTCTCCAAACCACCCCTT
CTTCACCGGACCCACCGCCGTCGCGCGCCAAACAACATCCATCTCTATCAAGCAGCTCC
TCCGCAGCAGCCACAAACATCTCCAGTTCTCCTCATCCATCTATTTCCACCCCTCCTTA
CTCTGACATGATTTGCACGGCGATTGCAGCGTTAAACGAACAGATGGGTCAAGCAAGCA
AGCTATTTTCGAGGTACATAGAGAGAATTTACACTGGGATTCCTACTGCTCATGGAGCTTT
GTTGACACACCATCTCAAGACTTTGAAGACCAGTGGGATTCCTTGTCTATGGTTAAGAAATC
TTACAAGCTTGCTTCTACTCCTCCTCCTCCTCCTCTACTAGTGTAGCTCCTAGTCTTGA
ACCTCCCAGATCTGATTTTCATAGTCAACGAGAACCAACCTTTACCTGATCCGGTTTTGGC
TTCTTCTACTCCTCAGACTATTAAACGTGGTCTGTTGGTGCACCTCCAAAGCTAAACCAGA
TGTTGTTCAACCTCAACCTCTGACTAATGGAAAACCTCACCTGGGAACAGAGTGAATTACC
TGTCTCTCGACCAGAGGAGATACAGATACAGCCGCCACAGTTACCGTTACAGCCACAGCA
GCCGGTTAAGAGACCGCGGGTCTGCTCTAGAAAAGATGGAACCTCGCCGACGGTGAAGCC
AGCTGCTTCTGTTTCCGGTGGTGTGGAGACTGTGAAACGAAGAGGTAGACCTCCGAGTGG
AAGAGCTGCTGGGAGGGAGAGAAAGCCTATAGTAGTCTCAGCTCCAGCTTCAGTGTTCCT
GTATGTTGCTAATGGTGGTGTGTAGACGCCGAGGGAGACCAAAGAGAGTTGACGCTGGTGG
TGCTTCTCTGTTGCTCCACCACCACCACCACCACTAACGTAGAGAGTGGAGGAGAGGA
GGTTGCAGTCAAGAAACGAGGAAGAGGACGGCCTCCTAAGATTGGAGGTGTATCAGGAA
GCCTATGAAGCCGATGAGAAGCTTTGCTCGTACTGGAACCCGTAGGAAGACCCAGAAA
GAATGCGGTGTCTAGTGGGAGCTTCTGACGACAAAGATGGTACTATGGAGAACTGAAGAA
GAAGTTTGAGTTGTTTCAAGCGAGAGCTAAGGATATTGTAATTGTGTTGAAATCCGAGAT
AGGAGGAAGTGGAAATCAAGCAGTGGTTCAAGCCATACAGGACCTGGAAGGGATAGCAGA
GACAACAAACGAGCCAAAGCACATGGAAGAAGTGCAGCTGCCAGACGAGGAACACCTTGA
AACCGAACCAGAAGCAGAGGGTCAAGGACAGACAGAAGCAGAGGCAATGCAAGAAGCTCT
GTTCTAAAGATAAAGCCTTGACATAAAAAGCTAGCAAGTGGTGGGTTTACTTGTGTGTG
TTACATGAAATTTTAAATCTTATAAGGGTGTGTCAGGAGAAAAACAAAAGAACAATGT
GATGAACTGATGATGATGATTGTGTCTCTAACCAACAACAAGGAGAGGTAGGGTAATGT
CTGTAAAGTGAATTAGGATGTTACCATTTGTTTCATGCTTCCCATCTCTCTCCATCGTCCAT
ATCTGTGTAGGCAGCTTTGTTCTTTGTTCCCTCGTGTGTTTTTTTAGACTGTTGTGTCTCT
TATTCTATTTTGTCTCCTTAGGCCTTTTAGGAGTTGTTGTTGATGTTTATCAAAAACGCT
TATGTAATTTTATGACCACTTCTACTTTTATGATGGTTTTCTT

>G1367 Amino Acid Sequence (domain in AA coordinates: 179-201, 262-285, 298-319, 335-357)

MDPSLSATNDPHHPPPPQFTSFPPFTNTNPFASPNHPFFTGPTAVAPPNNIHLYQAAPPQ
QPQTSVPVPHPSISHPPYSDMICTAIAALNEPDGSSKQAIISRYIERIYTGIPAHGALLT
HHLKTLKTSGLILVMVKSKYKLASTPPPPPTSVAPSLEPPRSDFIVNENQPLPDPVLASS
TPQTIKRGRRPFPKAKPDVVQPQPLTNGKLTWEQSELVSRPEEIQIQPPQLPLQPQPV
KRPPGRPRKDGTSPTVKPAASVSGGVETVKRRGRPPSGRAAGRERKPIVVSAPASVFPYV
ANGGVRRRRGRPKRVDAGGASSVAPPPPPPTNVESGGEEVAVKKRGRGRPPKIGGVIRKPM
KPMRSFARTGKPVGRPRKNAVSVGASGRQGDYGELEKKKFFELFQARAKDIVIVLKSEIGG
SGNQAVVQAIQDLEGIAETTNEPKHMEEVQLPDEEHLETEPEAEQGQTEAEAMQEALF*
>G1386 (89..673)

AATTTTATTTCTTCTCTCAAATCTTCCCACCAAAAATTAACCTTTTCGTTCACTAAG
TCCCTTTTAAAGAAAATATCCCAATTAATGGAACGTGACGACTGCCGGAGATTTTCAGGA
CTCGCCGGCGCAGACGACGAGAGAGAGTGAATATAAACCAAGAAGAAAAGAGCCAA
AGATGATGATGATGAGAAAAGTTGTTTCGAAGCATCCAAATTTTCGAGGTGTCAGAATGAG
ACAAATGGGGAAAATGGGTGTCCGAAATCAGAGAGCCAAAAAGAAATCAAGAATCTGGCT
CGGTACTTTTCTCACGGCGGAGATGGCGCGCGTGTCTACGACGTGGCAGCTTTAGCCAT
CAAAGGCGGTTCTGCACATCTCAACTTCCCGGAGCTCGCTTATCACCTCCCTAGACCAGC
TAGTGCCGACCCTAAAGACATCCAAGCTGCCGCGCGCGCAGCTGCAGCCGCTGTGGCCAT
TGACATGGATGTAGAGACGTCTTCGCCGTGCCATCTCCACAGTTACGGAAACGTCATC
TCCGCTATGATAGCACTCTCCGACGACGCGTTCTCCGATCTTCTGATCTCTTGCTCAA
CGTGAACCATAACATCGATGGCTTCTGGGACTCTTTTCCCTATGAAGAACCTTCTCTC
TCAAAGTTACTAGAACTCAAACATATGTCGTTTGTATGTATTTTGTATGTGACCA
TTTTTTGACGTCGAAAATCACCCGATAATCCAAATGTATGATTATTAATGGTTGATG
ATTTTCTTTGTGTGGAACAATGTGTATGATACGTAATCAAAAGTTCAAAAAAAAATAAA
AAAAA

>G1386 Amino Acid Sequence (domain in AA coordinates: TBD)
MERDDCRRFQDSPAQTERRVKYKPKKRAKDDDEKVVSKHPNFRGVRMRQWGWVSEI
REPKKKSRIWLGTFTAEMAARAHDAALAIKGSAGHLNFPDELAYHLRPPASADPKDIQA
AAAAAAAVIDMDVETSSPSPTVTETSSPAMIALSDDAFSDLPDLLNVNHNIDGF
DSFPYEPFLSQSY*

>G1421 (292..1155)

GAAATTTTCATCCCTAAATAAGAAAAAGCATCTCCTTCTTTAGTGTCTCTCTTACCAA
CTCTTGATTCCATAAGCATATATTAAGGCTCTCTGCTTTCTTCAACTTTCCCGGAA
AATCTTCTTGTACAAAGCATCAATCTCTTGTGTTTACCAATTTTCTCTTTATTCTTT
TTTGCCCTTTACTTTTCTTAACCTTTGGTCTTTATATATAAACACACGACACAAAGAAGAA
CACACATAAGTTAAACTATTACAACAGTTTTTAAAGAGAGAGATTTAAAAAATGGAGACA
GAGAAGAAAGTTTCTCTCCCAAGAACTTACGAATCTCTGTTACTGATCCTTACGCAACA
GATTCGTCAAGCGACGAAGAAGAAGATTGATTTGATGCATTATCTACAAAACGACGT
CGTGTTAAGAAGTACGTGAAGGAAGTGGTGTCTTGATTCGGTGGTTTCTGATAAAGAGAAG
CCGATGAAGAAGAAGAGAAAGAACGCGTGTGTTACTGTTCCAGTGGTTGTACGACGGCG
ACGAGGAAGTTTCTGAGGAGTGAAGCAAAGACCGTGGGGAAAATGGGCGGCGGAGATTAGA
GATCCGAGTAGACGTGTTAGGTTTGGTTAGGTACTTTTGACACGGCGGAGGAAGCTGCC
ATTGTTTACGATAACGCAGCTATTCAGCTACGTGGTCTAACGCAGAGCTTAACCTCCCT
CCTCTCCCGTGACGGAGAATGTGAAGAAGCTTCGACGGAGGTGAAAGGAGTTTCGGAT
TTTATCATTTGGCGGTGGAGAATGTCTTCGTTTCGCCGGTTTCTGTTCTCGAATCTCCGTT
TCCGGCGAGTCTACTGCGGTTAAAGAGGAGTTTGTGCGTGTATCGACGGCGGAGATTGTG
GTTAAAAAGGAGCCGTCTTTTACGGTTTCAAGATATTGGGGATTATTTCGGGTCGGATCCT
GACGTTTTTGTGTTTCTCTGACGTGAGTGAAAGTTTCGGCGGCGATTATTATTGGAGAT
AATCTTTTTCGGATATGAGTTTGGATCCGGGTTTGGATTTCGGGCTGGGCTGGATTTC
TCCAGCTGGCAGCTTGAGGACCATTTTCAAGATATTGGGGATTATTTCGGGTCGGATCCT
GTCTTAACGTGTTAAGAAATAACTGGCCGTTTAAACGGCGTTTAGTGAAGTTTGTGTTACCG
GCGACGGCGAGGATTAAAAAAAACGGCGATTATTTTTTTGAATGAAGATTGTGTTAAATA

>G1421 Amino Acid Sequence (domain in AA coordinates: 74-151)

METEKKVSLPRILRISVTDPYATDSSSDEEEVDFDALSTKRRRVKKYVKEVVLDVSVSD
KEKPMKKRKRKRVTVPVVVTTATRKFRGVRQRPWGKWAABIRDPSRRVRVWLGTFDTAE
EAAIVYDNAAIQLRGPNAELNFPFPPVTENVEEASTEVKGVSDFIIGGGECLRSPVSVLE
SPFSGESTAVKEEFVGVSTAEIVVKEPSFNGSDFSAFLPSDDVDFGFSTSMSESFGGDL
FGDNLFADMSFGSGFGSGSGFSSWHVEDHFDIGDLFGSDPVLTV*

>G1453 (39..917)

CGTCGACGCGAAATAAATCCTAGAAAATAACTATCAATATGATGAAGGTTGATCAAGATT
ATTCGTGTAGTATACCGCCTGGATTAGGTTTCATCCGACAGATGAAGAACTTGTTCGGAT
ATTATCTCAAGAAGAAAATCGCCTCCAGAGGATTGATCTCGACGTTATCAGAGAAATTG
ATCTTTACAAGATCGAACCATTGGGATCTACAAGAGAGATGTAGGATAGGGTACGAGGAGC
AAACGGAGTGGTATTTCTTACGCCATAGAGACAAGAAGTATCCGACTGGGACTAGGACAA
ACCGAGCCACCGTGCCGGTTTCTGAAAGCAACGGGCGGGACAAGGCGGTTTACCTCA
ACTCCAAACTTATCGGTATGAGAAAAACGCTTGTCTTTTACCGAGGTCGAGCGCTAATG
GCCAAAAGTCCGATTGGATCATTCACGAATACTACAGCTCGAGTCACACCAGAACTCTC

CTCCACAGGAAGAAGGATGGGTAGTGTGTAGAGCATTTAAGAAACGAACGACCATCCCCA
CAAAAAGGAGGCAACTTTGGGATCCGAAGTCTTATTTCTACGACGACGCCACTCTCTTGG
AACCTCTCGACAAGCGAGCCAGACATAATCCTGATTTTACCGCCACACCGTTCAAGCAAG
AACTACTCTCCGAGGCCAGTCACTGCCAGGATGGAGATTTTCGGATCTATGTACCTTCAAT
GCATCGATGATGATCAATTTCTCCAGCTTCTCAGCTCGAGAGCCCCCTCTCTCCGTCGG
AAATAACTCCCCATAGTACTACTTTTTCTGAGAACAGTAGCCGAAAGATGACATGAGCT
CCGAGAAGAGGATCACTGACTGGAGATATCTAGATAAGTTTCGTGGCGTCTCAATTTTGA
TGAGTGGAGAAGACTAAAAAAGGCTTTCCTATGCATGCATGCACTAGAAACGTTCGTCCGA
TTTTGGATTACATGCGGCCGCT

>G1453 Amino Acid Sequence (conserved domain in AA coordinates:13-160)
MMKVDQDYSCSIPPGRFHPDTDEELVGYYLKKKIASQRIDLDVIREIDLYKIEPWDLQER
CRIGYEEQTEWYFFSHRDKKYPTGTRTNRATVAGFWKATGRDKAVYLNLSKLIGMRKTLVF
YRGRAPNGQKSDWIIHEYYSLESHQNSPPQEEGWVVCRAFVKRRTTIPTKRRQLWDPNCLF
YDDATLLEPLDKRARHNPFTATPFKQELLSEASHVQDGFSGMYLQCIDDDQFSQLPQL
ESPSLPSEITPHSTTFSENSSRKDDMSSEKRITDWRYLDKFVASQFLMSGED*

>G1560 (120..1340)
ATCCTTTCAATTTCCACTCCTCTCTAATATAATTCACATTTTCCCACTATTGCTGATTCA
TTTTTTTTTGTGAATTATTTCAAACCCACATAAAAAAATCTTTGTTTAAATTTAAAACCA
TGGATCCTTCATTAGGTTTATTAAAGAGGAGTTTCTGCTGGATTGAGTATTCTCCAT
CACCACCATCTTCTTCTTATACCTTTATTCATCTTCCATGGCTGAAGCAGCCATAAATG
ATCCAACAACATTGAGCTATCCACAACCATTAGAAGGTCTCCATGAATCAGGGCCACCTC
CATTTTTGACAAAGACATATGACTTGGTGGAAGATTCAAGAACCAATCATGTCTGTCTT
GGAGCAAATCCAATAACAGCTTCATTGTCTGGGATCCACAGGCCTTTTCTGTAACCTCC
TTCCAGATTCTTCAAGCACAATAACTTCTCCAGTTTGTCCGCCAGCTCAACACATATG
GTTTCAGAAAGGTGAATCCGGATCGGTGGGAGTTTGCAAACGAAGGGTTTCTTAGAGGGC
AAAAGCATCTCCTCAAGAACATAAGGAGAAGAAAAACAAGTAATAATAGTAATCAAATGC
AACAACTCAAAGTTCTGAACAACAATCTCTAGACAATTTTTCATAGAAAGTGGGTAGGT
ACGGTCTAGATGGAGAGATGGACAGCCTAAGCGGAGACAAGCAAGTGTGATGATGGAGC
TAGTGAGACTAAGACAGCAACAACAAGCACCAAAATGTATCTCACATTGATTGAAGAGA
AGCTCAAGAAGACCGAGTCAAAACAACAACAATGATGAGCTTCTTGCCTCGCGCAATGC
AGAATCCAGATTTTATTGAGCAGCTAGTAGAGCAGAAGGAAAAGAGGAAAGAGATCGAAG
AGGCGATCAGCAAGAAGAGACAAAGACCGATCGATCAAGGAAAAGAAATGTGGAAGATT
ATGGTGATGAAAGTGGTTATGGGAATGATGTTGCAGCCTCATCCTCAGCATTGATTGGTA
TGAGTCAGGAATATACATATGGAAACATGTCTGAATTCGAGATGTCGGAGTTGGACAAAC
TTGCTATGCACATTCAAGGACTTGGAGATAATTCCAGTGCTAGGGAAGAAGTCTTGAATG
TGGAAAAGGAAATGATGAGGAAAGTAGAAGATCAACAACAAGGGTACCATAAGGAGA
ACAATGAGATTTATGGTGAAGGTTTTTGGGAAGATTTGTTAAATGAAGGTCAAAATTTTG
ATTTTGAAGGAGATCAAGAAAATGTTGATGTGTTAATTCAGCAACTTGGTTATTTGGGTT
CTAGTTCACACACTAATTAAGAAGAAATTGAAATGATGACTACTTTAAGCATTTGAATCA
ACTTGTTCCTATTAGTAATTTGGCTTTGTTTCAATCAAGTGAGTCGTGGACTAAGTGC

>G1560 Amino Acid Sequence (domain in AA coordinates: 62-151)
MDPSRFIKEEFPAGFSDSPSPSSSSYLYSSMAEAAINDPTTSLYPQPLEGLHESGPP
PFLTCTYDLVEDSRTNHVSVWSKSNNSFIVWDPQAFSVTLTLPFFKHNNFSSFVRQLNTY
GFRKVNPDWRWEFANEGFLRGQKHLKNI RRRKTSNNSNQMQPQSSEQQSLDNFCIEVGR
YGLDGEMLSLRRDKQVLMELVRLRQQQSTKMYLTLIEEKLKKTESKQKQMMFLARAM
QNPDFIQQLVEQKEKRKEIEEAI SKKRQRPIDQKRNVEDYDGDESGYGNDVAASSSALIG
MSQEYTYGNMSEFEMSELDKLAHMQGLGDNSSAREEVLNVEKGNDEEEVEDQQQGYHKE
NNEIYGEFWEDLLNEGQNFDFEGDQENVDLIQQLGYLGSSSHTN*

>G1594 (1..984)
ATGGATGGAATGTACAATTTCCATTCCGCCGGTGATTATTAGATAAGTCGGTTCTGATG
ATGTCACCGGAGAGTCTCATGTTTCTTCCGATTACCAAGCTTTGCTATGTTCTCCGCC
GGTGAAAATCGTGTCTCTGATGTTTTCCGGATCCGACGAGCTACTCTCAGTAGCCGTCTCC
GCTTTGTCGTCGGAGGCGGCTTCGATCGCTCCGAGATCCGAAGAAATGATGATAACGTT
TCTCTAAGTGTCAAAAGCTAAAATCGCTTGTCTATCTTTCGTATCCTCGCTTACTTCAA
GCTTACATCGATTGCCAAAAGGTCCGAGCACCACCGGAGATAGCGTGTTTACTAGAGGAG
ATTCAACGGGAGAGTGATGTTTATAAGCAAGAGGTTGTTTCTTCTTCTGCTTTGGAGCT
GATCCTGAGCTTGATGAATTTATGGAAACGTACTGCGATATATTAGTGAATACAAATCG

GATCTAGCAAGACCGTTTGACGAGGCAACGTGTTTCTTGAACAAGATTGAGATGCAGCTA
CGGAACCTATGTACTGGTGTGAGTCTGCCAGGGGAGTTTCTGAGGATGGTGTAATATCA
TCTGACGAGGAACCTGAGTGGAGGTGATCATGAGGTAGCAGAGGATGGGAGACAAAGATGT
GAAGACCGGGACCTCAAAGATAGGTTGCTACGCAAATTTGGAAGCCGTATTAGTACTTTA
AAGCTTGAGTTCTCAAAGAAGAAGAAGAAAGGAAAGTTACCAAGAGAAGCAAGACAAGCT
CTTCTTGATTGGTGGAACTCTCCATTATAAGTGGCCTTACCCTACTGAAGGAGATAAGATA
GCATTAGCTGATGCAACGGGGTTAGACCAAAAACAAATCAACAATTGGTTTATAAACC
AGGAAACGTCATTGGAAGCCATCAGAGAATATGCCTTTCGCTATGATGGATGATTCTAGT
GGATCATTTCTTTACCGAGGAATGA

>G1594 Amino Acid Sequence (conserved domain in AA coordinates: 343-308)

MDGMYNFHSAGDYSKSVLMMSPESLMFSPDYQALLCSSAGENRVSDVFGSDELLSVAVS
ALSSBAASIAPETIRNDDNVS LTVIKAKIACHPSYPRLLQAYIDCQKVGAPPEIACLL
IQRESDVYKQEVVPSSCFGADPELDEFMETYCDILVKYKSDLARPFDEATCFLNKIE
MQLRNLCTGVESARGVSESGVIS SDEELSGGDHEVAEDGRQRCEDRDLKDRLLRKFG
SRISTL KLEFSKKKKKGLPREARQALLDWN LHYKWPYPTEGDKIALADATGLDQKQIN
NWFINQ RKRHWKPSENMPFAMDDSSGSFFTEE*

>G1750 (94..1101)

CCCTTTTCTCTCTTTCTCCAAATCTCTGAAAATTTTCACCAGAATCTCTGTTCTTTT
TCACCAGAATCTCTCTGTTTAAAATAATAGGTGATGATGATGGATGAGTTTATGGATCT
AGACCAGTGAAGTACACAGAGCAAGACTGTTATCAGAAAGTACACTAAAAGTCGTCT
ATGGAGAGGAAGACCACTGTTCTGACTCGGCCAGGTTGGTTTCGGGTCTCAATGACGG
ATCGTGACGCCACTGATTCATCAAGCGACGAGGAAGAGTTTCTGTTCCCTCGAAGACG
TGTC AAGAGATTGATTAACGAGATCAGAGTCGAGCCTAGCAGCTCTCCACCGGCGACG
TCTCT GCTTCTCCGACGAAGGACCGGAAAAGAATCAACGTTGATTCTACGGTTCAA
AGCCCTCT GTTTCCGCCAAAACCAGAAGAGTACCGCGCGCTGAGACAGCGACCATG
GGGAAAATGG GCGGCGGAGATTCTGATCCTGAGCAACGCCGAGAAATCTGGCTCGG
TACTTTTGCAACG GCGGAGGAAGCTGCCATCGTCTACGACAACGCAGCAATCAA
ACTTCGTGGCCCTGATGCT CTTACCACTTCACCGTACAACCAAGAACCGGTACA
AGAACAAGAACCG GAGAGCAACATGTCGGTTTCGATATCAGAATCAATGGACGAT
TCTCAACATCTATCT CCGACATCGGTTCTCACTACCAACATATGTCTCGGAGGA
ACCAATCGATAGTCTTATC AAACCGGTTAAACAAGAGTTTCTTGAACCAGAACA
AGAGCCAATAAGCTGGCATCTTGGA GAAGGTAATACTAATACTAATGATGATT
CATTTCCATTGGACATTACATTTCTCGACAAC TATTTCAATGAATCATTACCAG
ACATCTCCATCTTCGATCAACCTATGTCTCTATTCAA CCAACAGAGAATGATTT
CTTCAACGACCTTATGTTATTCGATAGCAACGCAGAAGATAC TACTCCTCCG
AGATCAAAGAGATTGGTTCATCGTTCAACGATCTTGATGATTCTTTGATA
TCCGATCTCTTACTGTGTGATATTTTGGCATTAAACAAACACCGGTTTGGTTGC

>G1750 Amino Acid Sequence (domain in AA coordinates: 107-173)

MMDEFMDLRPVKYTEHKT VIRKYTKSSMERKTSVRDSARLVRVSMTRD
ATDSSSDEE EFLFPRRRVKRLINEIRVEPSSSTGDVSASPTKDRK
RINVDSTVQKPSVSGQNQKKYRG VRQRPWGKWAEBIRDPEQRRRI
WLGTAFATAEEAAI VYDNAAIKLRGPDALTNFTVQPEPE PVQE
EQEPEPNMSVSI SESMDDSQHLSSPTSVLNYQTYVSEEPIDSLIK
PVKQEFLEPE QEPISWHLGEGNTNTNDDSFPLDITFLDNYFNESL
PDISIFDQPMSP IQPTENDFFNDLM LFDSNAEEYYSSEIKEIGSS
FNDLDDSLISDLLLV*

>G1947 (70..918)

ACAACATTTCTCTCTCTCTTTTATTAAAAAGCTCAAATTTATATAGGTTTTTT
GTTCAAAAAATGGATTATAACCTTCCAATTCCATTAGAGGGTCTCAAAGAAACGCCACCA
ACGGCTTTCTTGACGAAAACATACAACATAGTGGAGGATTCAAGCACAAACATAGTT
TCATGGAGCAGAGACAACAACAGCTTCATTGTTTGGGAACCAGAGACTTTTGCCCTAATT
TGCCTCCCTAGATGCTTTAAGCACAAATAATTTCTCCAGCTTTGTTAGACAGCTCAACT
TATGGGTTTAAAGAAGATTGATACAGAGAGATGGGAATTTGCAAATGAGCATTTTCTGAAG
GGAGAGAGGCATCTTCTTAAGAACATCAAGAGAAGAAGACATCATCTCAAACGCAAACG
CAGTCGCTAGAAGGAGAGATCCATGAGCTGCGAAGAGACAGAATGGCTTTAGAAGTAGAA
CTGGTTAGACTGCGCAGAAAACAAGAAAGCGTGAAGACATATCTGCATTTGATGGAAGAG
AAACTGAAAGTCAAGAGTAAAGCAAGAAATGATGATGAATTTCTTGCTAAAGAAGATT
AAGAAACCGAGTTTTTACAGAGCTTAAGGAAACGTAATCTGCAAGGAATCAAGAATCGA
GAGCAAAAGCAAGAGGTGATCTCAAGCCATGGTGTGAGGATAATGGAAAGTTTGTAA
AAGCTGAGCCAGAAGAGATGGTGTGATGACATCGATGATCAATGTGGAGGTGTGTTTGATTAT

GGTGATGAGCTTCACATAGCTTCAATGGAGCATCAAGGACAAGGGGAGGATGAAATTGAA
ATGGATAGTGAAGGAATTTGGAAGGGTTTCGTGTTGAGTGAGGAGGAGATGTGTGATTTA
GTGGAACATTTTATATAATAAACTAATGTATTATGAGAGGTTTTTTTTTGTGTTTTTGCT
TTTTTTTTCCGAGTTTGTCAAGCATTGTATACAATTTGGGCCAAACTAAAAGCCCAA
CAAAATATTTGGCCTTGGCATTGTGTTAACAAATTGACTAATTCGCCACACCTTCC
>G1947 Amino Acid Sequence (domain in AA coordinates: 37-120)
MDYNLPIPLEGLKETPPTAFLTKTYNIVEDSSTNNIVSWSRDNSFVWEPETFALICLP
RCFKHNNFSSFVRQLNTYGFKKIDTERWEFANEHFLKGERHLLKNIKRRKTSSQTQTQSL
EGEIHLELRDRMALEVELVRLRRKQESVKTYLHLMEEKLKVTEVKQEMMMNFLKKIKKP
SPLQSLRKRNQLGQIKNREKQKQEVSSHGVEDNGKFVKAEPPEYGDIDDDQCGGVFDYDGE
LHIASMEHQGGGEDEIEMDSEGIWKGFVLSEEMCDLVEHFI*
>G2011 (309..1547)
AATGTCGGTTGTACAATTATTTGTCACTAAAGTTTCCAAATTTCTTCTAAACTGATGAAT
CAATGGAACATGATGACGAAAAAGATAAATCCACGGTGGCGGGAACCTGACCCACCCATTT
CCACCGCTCTCTATTTCCCCAGATTTTTTTCAATTATCTGACTACAGTTTGTGCGTTACT
TCCTTCCCTAAACCTTTATAAACCATTAAACCTCTCATCTTCTTCTTAAACCCCTA
ATTATCACACACACCCCAATTTCTCACTCTCTCTCTCACTAAAACCCGTAAATTTTCTAC
TATATCAAATGAGCCCCAAAAAAGATGCTGTTTTCTAAACCAACTCCAATTTAGTACCCG
TTTTCGAGACGATCCGATATACCCGGGTCTCTCTACGTGACACTGACATGGGTTTTCTCTG
GGTCACCACTTCCCATGCCACTAGACATCTTACAAGGGAATCCAATTCACCTTTTTTAT
CCAAGACTTTTGATTGTTGTTGATGACCCGACTCTTGACCCGGTCATCTTTGGGGACTGA
CCGGAGCTAGCTTCGTAGTTTGGGATCCTCTAGAGTTTGCCAGAATCATACTTCCAAGGA
ATTTCAAACACAACAATTTCTCCAGCTTCGTGACAGCTTAACACTTATGGATTTTCGAA
AGATTGATACTGACAAGTGGGAATTCGCTAACGAGGCTTCTCTTAGAGGCAAGAAGCATC
TTCTGAAGAACATTATCGTCGTGATCACCACAATCCAACCAAACTTGCTGCAGTAGCA
CTAGCCAAAGCCAAAGGGTCACTACTGAGGTTGGAGGAGAGATTGAGAAGCTGAGGAAAG
AGCGGCGTGCATTGATGGAGGAAATGGTTGAGCTTCAGCAGCAAAGCAGAGGCACAGCTC
GACATGTGGACACTGTAAACCAGAGGCTGAAAGCTGCAGAGCAACGTCAGAAGCAATTGC
TCTCTTTCTTGGCTAAGTTGTTTCAGAACCGGGTTTCTTGGAACGCCTGAAGAATTCA
AAGGAAAAGAAAAGGAGGAGCTCTTGATTGGAAGGCGAGAAAGAAAGTTTCATCAAGC
ACCACCAGCAGCCTCAAGATTCTCCAACAGGAGGGGAGGTGGTGAAGTATGAAGCTGATG
ATTGGGAGAGATTGCTAATGTATGACGAAGAGACTGAGAACACCAAGGGTTTAGGAGGA
TGACTTCAAGCGATCCAAAAGGCAAGAACTTGATGTATCCATCAGAAGAAGAGATGAGCA
AACCAGATTACTTGATGCTCTCCCATCTCCTGAAGGACTTATTAACAAGAAGAGACGA
CATGGAGCATGGGTTTCGATACTACAATACCGAGTTTCAGCAACACCGATGCATGGGGAA
ACACAATGGACTATAATGATGTCTCAGAGTTTGGTTTTGCTGCAGAAACAACAAGTGATG
GTTTGCCGTGATGTCTGCTGGGAACAATTTGCTGCAGGAATCAGAGACTGGATTCAACT
GGCCAACTGGTGATGATGATGATAATACGCCAATGAATGATCCTTAGGATCTTTTCATAT
ATAGTTTATAGACAAAACCCGTTTCTTATCGGGTGAACATTATTAATTCATTATTCATTTG
AATGCACTCTTTATACATATATATAATATTGATGAGTTTGATTGTTCCAAAAA
>G2011 Amino Acid Sequence (domain in AA coordinates: 56-147)
MSPKKDAVSKPTPIPVSVRRSDIPGSLYVDTDMGFSGLPLPMLDILQGNPIPPFLSKT
FDLVDDPTLDPVISWGLTGASFVVDPLEFARIILPRNFKHNNFSSFVRQLNTYGFRRKID
TDKWEFANEAFLRGKKHLLKNIHRRRSPQSNQTCSSSTSQSQSPTEVGGIEKLRKERR
ALMEEMVELQQQSRGTARHVDTVNQRLLKAAEQRQKQLLSFLAKLFQNRGFLERLKNFKGK
EKGGALGLEKARKKFIKHHQPPQDSPTGGEVVKYEADDWERLLMYDEETENTKGLGGMTS
SDPKGKNLMPSEEMSKPDYLSFSPSEGLIKQEETWMSGFDTTIPSFSNTDAWGNMT
DYNDVSEFGFAAETTS DGLPDVCWEQFAAGITETGFNWPTGDDDDNTPMNDP*
>G2094 (1..450)
ATGCTAGATCCACCGAGAAAGTAATCGATTGAGAATCAATGGAAAGCAAACCTCACATCA
GTAGATGCGATCGAAGAACACAGCAGCAGTAGCAGTAATGAAGCTATCAGCAACGAGAAG
AAGAGTTGTGCCATTTGTGGTACCAGCAAAACCCCTCTTTGGCGAGGCGGTCTGCGGT
CCCAAGTCGCTTTGTAACGCATGCGGGATCAGAAACAGAAAGAAAAGAAGAACTGATC
TCAAATAGATCAGAAGATAAGAAGAAGAGTCATAACAGAAACCCGAAGTTTGGTGAC
TCGTTGAAGCAGCGATTAATGGAATTGGCGAGAGAAGTGATGATGCAGCGATCAACGGCT
GAGAATCAACGGCGGAATAAGCTTGGCGAAGAAGAGCAAGCCGCGGTGTTACTCATGGCT
CTCTCTTATGCTTCTTCCGTTTATGCTTAA

>G2094 Amino Acid Sequence (domain in AA coordinates:43-68)
MLDPTEKVIDSESMESKLTSDVAIEEHSSSSSNEAISNEKKSCAICGTSKTPLRGGPAG
PKSLCNACGIRNRKKRRTLISNRSEDKKKKSHNRNPKFGDSLKQRLMELGREVMMQRSTA
ENQRRNKLGEEEQAAVLLMALSYASSVYA*

>G2113 (90..590)

ATAACAACTCATCAAACCTCCCTCAGCGTTTCTTTTTCTTACATAAACAATTTTTCTTAC
ATAAACAAATCTTGTGTTTGTGTTGTGTCATGGCACCAGACAGTTAAACGGCGGCCGTCA
AAACCAACGAAGGTAACGGAGTCCGTTACAGAGGAGTGAGGAAGAGACCATGGGGACGTT
ACGCAGCCGAGATCAGAGATCCCTTCAAGAAGTCACGTGTCTGGCTCGGTACTTTCGACA
CTCCTGAAGAAGCCGCTCGTGCCCTACGACAAACGTGCTATTGAGTTTCGTGGAGCTAAAG
CCAAAACCAACTTCCCTTGTTACAACATCAACGCCCACTGCTTGAGTTTGACACAGAGCC
TGAGCCAGAGCAGCACCGTGGAATCATCGTTTCCCTAATCTCAACCTCGGATCTGACTCTG
TTAGTTTCGAGATTCCCTTTTCCCTAAGATTCAAGTTAAGGCTGGGATGATGGTGTTCGATG
AAAGGAGTGAATCGGATTCTTCGTCGGTGGTGTGATGGATGTCGTTAGATATGAAGGACGAC
GTGTGGTTTTGGACTTGGATCTTAATTTCCCTCCTCCACCTGAGAACTGATTAAGATTTA
ATTATGATTATTAGATATAATTAATGTTTCTGAATTGAG

>G2113 Amino Acid Sequence (domain in AA coordinates: TBD)

MAPTVKTAAVKTNEGNGVRYRGVRKRPWGRYAAEIRDPFKKSRVWLGTFTDPEEAARAYD
KRAIEFRGAKAKTNFPYCYNINAHCLSLTQSLSQSSSTVESSFNPNLNLGSDSVSSRFPFKI
QVKAGMMVFDERSSESDSSSVMDVVRYEGRRVLDLNLNFPFPPEN*

>G2115 (41..733)

AATCACTCTACAAAGCCTGTACGTACACAACAACATTACCATGGTGAAACAAGAACGCAA
GATCCAAACCAGCAGCACAAAAAGGAAATGCCTTTGTTCATCATCACCATCTTCTTCTTC
TTCTTCATCTTCTTCTCCTCGTCTTCGTCTTCGTGTAAGAACAAGAACAAGAAGAGTAAGAT
TAAGAAGTACAAAGGAGTGAGGATGAGAAGTTGGGGATCATGGGTCTCTGAGATTAGGGC
ACCAATCAAAAGACAAGGATTTGGTTAGGTTCTTACTCAACAGCTGAAGCAGCTGCTAG
AGCTTACGATGTTGCACTCTTATGTCTCAAAGGCCCTCAAGCCAATCTCAACTTCCCTAC
TTCTTCTTCTTCTCATCATCTTCTTGATAATCTCTTAGATGAAATACCTTTTGTCCCC
CAAATCCATCCAAAGAGTAGCTGCTCAAGCTGCCAACTCATTTAACCATTTTGCCCTAC
TTCATCAGCCGTCTCGTCACCGTCCGATCATGATCATCACCATGATGATGGGATGCAATC
TTTGATGGGATCTTTTGTGGACAATCATGTGTCTTTGATGGATTCAACATCTTCATGGTA
TGATGATCATAATGGGATGTTCTTGTGTTGATAATGGAGTCCATTCAATTACTCTCCTCA
ACTAACTCGACGACGATGCTCGATGAATACTTCTACGAAGATGCTGACATTCCGCTTTG
GAGTTTCAATTAATCCGACGGTCCATAATACATACTTTAATTAGT

>G2115 Amino Acid Sequence (conserved domain in AA coordinates:46-115)

MVKQERKIQTSSTKEMPLSSSSSSSSSSSSSSCKNKNKSKIKKYKGVMRMSWGS
WVSEIRAPNQKTRIWLGSYSTAEAAARAYDVALLCLKGPQANLNFPTSSSSHLLDNLLD
ENTLLSPKSIQRVAAQAANSFNHFAPTSSAVSSPSDHDHHDGMSLMGSFVDNHVSLM
DSTSSWYDDHNGMFLFDNGAPFNYSPLNSTTMLDEYFYEDADIPLWSFN*

>G2130 (41..988)

CCTCTCTTCATTTTTTAACCTCCCTCTCTCTCTCTCTCTATGGAGAGACGAACGAGACG
AGTGAAGTTCACAGAGAATCGTACGGTCACAAACGTAGCAGCTACACCATCTAACGGGTC
TCCGAGACTGGTCCGTATCACTGTACTGATCCTTTTCGCTACTGACTCGTCTAGCGACGA
CGACGACAAACAACGTACCGTGTTCCAAAGAGTGAAACGATACGTGAAGGAGATTAG
ATTCTGCCAAGGTGAATCTTCTTCTCCACCGCGGCGAGGAAAGGTAAGCACAAAGGAGGA
GGAAAGCGTAGTGGTTGAAGATGACGTGTGACGTCGGTGAAGCCTAAAAAGTACAGAGG
CGTGAGACAGAGACCTTGGGGAAAATTGCGGCGGAGATTAGAGATCCGTCGAGCCGTAC
TCGGATTTGGCTTGGGACTTTTGTACGGCGGAGGAAGCTGCTATAGCGTACGATAGAGC
CGCGATTCACTCAAAGGACCTAAAGCGCTCACGAATTTCTTAACCTCCGCCGACGCCAAC
GCCGGTTATCGATCTCAAACGGTTTCCGCCTGCGATTACGGTAGAGATTCTCGGCAGAG
CCTTCATTACCGACCTCTGTTCTAAGATTCAACGTCAACGAGGAAACAGAGCATGAGAT
TGAAGCGATCGAGCTATCTCCGAGAGAGAAAGTCGACGGTTATAAAAGAAGAAGAAGAATC
GTCGGCGGGTTTGGTGTTCCTCGATCCGTATCTGTTACCGGATTTATCTCTCGCCGGCGA
ATGTTTTTGGGATACCGAAATTGCCCTGACCTTTTGTCTCGATGAAGAAACCAAAT
CCAATCAACGTTGTTACCAAACACAGAGGTTTCGAAACAAGGAGAAAACGAACTGAAGA
TTTCGAGTTTGGTTTGATTGATGATTTTCGAGTCTTCTCCATGGGATGTGGATCATTTCTT
CGACCATCATCATCACTCTTTCGATTAAAAATCTCTTCTTTTTTGGGGAAATTTTTGTG

>G2130 Amino Acid Sequence (domain in AA coordinates 93-160)
MERRTRRVKFTENRTVTNVAATPSNGSPRLVRITVTDPFATDSSSDDDNNVTVVPRVK
RYVKEIRFCQGESSSSTAARKGKHKEESVVVEDDVSTSVKPKYRGVRQRPWGKFAAEI
RDPSSRTRIWLGTFTVTAEEAAIAYDRAAIHLKGPKALTNFLTPPTPTVIDLQTVSACDY
GRDSRQSLHSPTSVLRFNVNEETEHEIEAIELSPERKSTVIKEEEESSAGLVFPDPYLLP
DLSLAGECFWDTEIAPDLLFLDEETKIQSTLLPNTEVSKQGENETEDFEFGLIDDFESSP
WDVDHFFDHHHSFD*

>G2147 (162..1262)

CTGTGATGTGCAAGAGTTTGAACACACAAAGAAGAAAGAAGAACTCAACATTTCAAGCAA
GAAGAAAGAGAGAAGAGAGAAGGTCCAATAATAGAGAGAACAAAAAAGAGAGCTTAA
TTGTCAAGTTTATTCTCTGCAACCGTGGCGCCTAAGTAACACATGTGCAATTATGGAGTTA
AAGAGCTCACATGGGAAAAATGGGCAACTAACCGTTTCATGGTCTAGGCGACGAAGTAGAAC
CAACCACTCGAATAACCCCTATTTGGACTCAAAGTCTCAACGGTTGTGAGACTTTGGAGT
CTGTGGTTCATCAAGCGGCTCTACAGCAGCCAAGCAAGTTTCAGCTGCAGAGTCCGAATG
GTCCAAACCACAATTATGAGAGCAAGGATGGATCTTGTTCAGAAAACGCGTTATCCTC
AAGAAATGGACCGATGGTTCGCTGTTCAGAGGAGAGCCATAGAGTTGGCCACAGCGTCA
CTGCAAGTGCGAGTGGTACCAATATGTCTTGGGCGTCTTTTGAATCCGGTGGAGCTTGA
AGACAGCTAGAACCAGGAGACAGAGACTATTTCCGCTCTGGATCGGAACTCAAGATACTG
AAGGAGATGAACAAGAGACAAGAGGAGAAGCAGGTAGATCTAATGGACGACGGGGACGAG
CAGCAGCGATTACAAACGAGTCCGAAAGGAGACGGCGTGATAGGATAAACAGAGGATGA
GAACACTTCAGAAGCTGCTTCTTCTACTGCAAGTAAGGCGGATAAAGTCTCAATCTTGGATG
ATGTTATCGAACACTTGAAACAGCTACAAGCACAAAGTACAGTTTCATGAGCCTAAGAGCCA
ACTTGCCACAACAATGATGATTCCGCAACTACCTCCACCACAGTCAGTTCTCAGCATCC
AACACCAACAACAACAACAACAACAGCAGCAGCAGCAGCAACAACAGCAGCAACAGTTTC
AGATGTGCTGTGCTTGAACAATGGCAAGAATGGGAATGGGAGGTGGTGGAAATGGTTATG
GAGGTTTGTGTTCTCTCTCTCTCTCTCTCCACCAATGATGGTCCCTCTATGGGTAACAGAG
ACTGCACCAACGGTTCTTCAGCCACATTATCTGATCCATACAGCGCTTTTTCGCACAGA
CAATGAATATGGATCTCTACAATAAAATGGCAGCAGCTATCTATAGACAACAGTCTGATC
AAACAACAAGGTAAATATCGGCATGCTTCAAGTTCTTGAATCATGAGAAAAGAGATT
AGTCTAGCGACCTAGTATTATTGATCCATATATAGTTCTTGAAGATTGTTGTATCAT
GATTGTAAAAACTGTTTTGAGTATGGAAAAAGACTTGCAGATAAAA

>G2147 Amino Acid Sequence (domain in AA coordinates:160-234)
MSNYGVKELTWENGQLTVHGLGDEVEPTTSNNPIIWTQSLNGCETLESVVHQAALQQPSKF
QLQSPNGPNHNYESKDGSCSRKRGYPQEMDRWFAVQEESHVGHVSVTASASGTNMSWASF
ESGRSLKTARTGDRDYFRSGSETQDTEGDEQETRGEAGRSNGRRGRAAAIHNESERRRRD
RINQRMRLTQKLPTASKADKVSILDDVIEHLKQLQAQVQFMSLRANLPQQMMIPQLPPP
QSVLSIQHQQQQQQQQQQQQFQMSLLATMARMGMGGGGNGYGGGLVPPPPPPPMV
PPMGNRDCITNGSSATLSDPYSAFFAQTMNMDLYNKMAAAIYRQSDQTTKVNIGMPSSSS
NHEKRD*

>G2156 (384..1292)

TTTTTTTTTCCCTTTCCTCGTTCAAAAAAGTACTTGCAGAGTCACTCACTCTCAGTCTCA
GCACATGAATTAATTTGAAGCTTCCCTAGAATTCTTTCACATCAATTAATACGACACCGT
CTCGGGTGAAGAATCTCTCTCTCTCTTGGCCCTAAAGCGAGTTAGGGTTTAACACACAAAGC
ATACCTTTTAGATTTGTGTCTCTTAGCTCTGTTTTTGTGCGCTTGTGTAAACCGATCAACT
CAAGCTATTGGCTCCTCACCTCCTGAAATTTGACTTCTCCAATGGATCTCAAAGTTTCTC
TTATATGAATTCATCTTACCCTCACAATATCTTTATATATATAGCCACAAGAACAAG
AAGAGTCAGTAGATGCGGCTGCCATGGACGGTGGTTACGATCAATCCGAGGAGCTTCTA
GATACTTTCACAACCTCTTCAGGCCTGAGCTTCATCACCAGCTTCAACCTCAGCCTCAAC
TTCACCCTTTGCTCAGCCTCAGCCTCAACCTCAGCCTCAGCAGCAGAAATTCAGATGATG
AATCTGACTCCAACAAGGATCCGGGTTCGACCCAGTTACCTCTGGTTCAACCGGGAAAC
GTCCACGTGGACGTCTCCGGGATCCAAGAACAAGCCGAAGCCACCGGTGATAGTACTA
GAGATAGCCCCAACGTGCTTAGATCTCATGTTCTTGAAGTCTCATCTGGAGCCGACATAG
TCGAGAGCGTTTACCCTTACGCTCGCAGGAGAGGAAGAGGAGTCTCCATTCTCAGTGGTA
ACGGCACGGTGGCTAACGTCACTCTCCGGCAGCCGGCAACGACAGCGGCTCATGGGGCAA
ATGGTGAACCGGAGGTGTTGTGGCTCTACATGGAAGGTTTGAAGATACTTTCCCTCACAG
GTACGGTGTGTCGCCCCCTGCGCCGCCAGGATCCGGTGGTCTTTCTATCTTTCTTCCG
GCGTTCAAGTCAAGGTGATTGGAGGAAACGTGGTGGCTCCGCTTGTGGCTTCGGGTCCAG

TGATACTAATGGCTGCATCGTTCTCTAATGCAACTTTTCGAAAGGCTTCCCCTTGAAGATG
AAGGAGGAGAAGGTGGAGAGGGAGGAGAAGTTGGAGAGGGAGGAGGAGAAGGTGGTC
CACCGCCGGCCACGTCATCATCACCACCATCTGGAGCCGGTCAAGGACAGTTAAGAGGTA
ACATGAGTGGTTATGATCAGTTTGCCGGTGATCCTCATTGCTTGGTTGGGAGCCGCAG
CCGCAGCCGCACCACCAAGACCAGCCTTTTAGAATTGAAAATTATGTCCGTAACATAGCT
GTAACCAAATTTCAATTTCTCAAAATTAAGAAAAA

>G2156 Amino Acid Sequence (domain in AA coordinates:66-86)
MDGGYDQSGGASRYFHNLFPELHHQLQPQQLHPLPQPQPQPQPNQNSDDESNSNDP
GSDPVTSGSTGKRPRGRPPGSKNPKPPVIVTRDSPNVLRSHVLEVSSGADIVESVTTYA
RRRGRGVSILSGNGTVANVSLRQPATTAHANGGTTGGVVALHGRFEILSLTGTVLPPPA
PPSGGLSIFLSGVQGVIGNVVAPLVASGPVILMAASFNATFERLPLEDEGGEGGEG
GEVGEgggggEGGPPATSSSPSGAGQGQLRGNMSGYDQFAGDPHLLGWGAAAAAAPP RP
AF*

>G2294 (24..659)

TCCTCCCTTAATTAGTATCAAAAATGGTGAAAACACTTCAAAAGACACCAAAGAGAATGT
CATCTCCATCATCATCATCTTCATCATCCTCATCAACATCATCATCATCCATAAGGATGA
AGAAGTACAAGGGAGTGAGAATGAGAAGTTGGGGTTTCATGGGTTTCAGAGATCAGAGCTC
CTAATCAAAAGACAAGGATCTGGCTTGGTTCTTACTCAACTGCTGAAGCCGCGGCTAGAG
CCTACGACGCAGCACTCTATGTCTTAAAGGATCCTCAGCTAATAATCTCAACTTCCCAG
AGATCTCAACTTCTCTTACCATATTATCAACAATGGTGATAACAACAATGACATGTCCC
CTAAGTCTATACAAGAGTAGCAGCTGCAGCTGCTGCTGCCAACACAGATCCTTCTCAT
CATCAGTCTCTACTTCTCTCCATTGCTTTCCTCTCCATCTGAAGATCTCTATGATGTTG
TCTCCATGTACAGTATGACCAACAAGTCTCCTTGTCTGAATCATCATCATGGTACAAC
GCTTTGATGGTGATGATCAGTTTATGTTTCAATATGGAGTCTCCGCGCCGTATTTGACAA
CATCACTTTCTGATGATTTCTTTGAGGAAGGAGATATCAGATTATGGAACCTCTGCTGAT
TCTACTTTTCAATTATACCTTATTCCTTTG

>G2294 Amino Acid Sequence (conserved domain in AA coordinates:32-102)

MVKTLQKTPKRMSSPSSSSSSSSSSSSSSSSSIRMKKYKGVMRWSWGSVWSEIRAPNQKTRIW
LGSYSTAEAAARAYDAALLCLKGSSANNLNFPEISTSLYHIINNGDNNNDMSPKSIQVVA
AAAAAANTDPSSSSSVTSSPLLSSPSEDLYDVVSMSQYDQVSLSESSSWYNCFDGDDQF
MFINGVSAPYLTTSLSDDFEEDIRLWNFC*

>G2510 (16..594)

ATAACAAACTCTTTAATGTCAACACAGAGAATGAAGCTATCATCACCACCAGTTACCAAC
AACGAACCAACCCGCCACCGCTTCTGCCGTTAAATCTTGCGGCGGAGGAGGTAAAGAAACC
AGCTCATCGACACACGAGGCATCCAGTGATACCACGGAGTTTCGCAAACGCCGATGGGGAAAA
TGGGTTTCTGAGATCAGAGAGCCCCGGAAGAGTCTCGGATTTGGCTCGGATCTTTTCCG
GTGCCGGAGATGGCTGCTAAGGCCTACGACGTGGCAGCGTTTTGTCTAAAAGGTAGAAAA
GCTCAGCTGAATTTCCCTGAAGAAATCGAGGATCTACCTCGACCGTCCACGTGTACTCCC
AGAGATATCCAAGTCGACGCGCCAAAGCAGCCAACGCCGTGAAGATCATCAAAATGGGA
GATGATGACGTGGCAGGAATAGACGACGGAGATGATTTCTGGGAAGGCATTGAGCTGCCT
GAGCTTATGATGAGTGGAGGTGGGTGGTTCGCCGAGCCTTTTGTTCGCCGAGATGATGCC
ACGTGGCTTGTTCGACGGAGACTTGTATCAGTATCAGTTTATGCGGTGTCTGTGAGTGTG
CTGTTCGATTGTGTCTGATTCTGTTATACGTGTACGTGTATGTTTGTGTTGGCTCACT
TAATTTAATGCATATGCATGTATATTTTCAATTTATTTGTTTCTAGTTTATGTTTTACGC
GATTAATAATTAGATACCTGTTTCTCAAGTTAGTTATCAGGTTTGTACGCATCTACAAAA
ATACGTATAAGTGTATGTTCTTATATACAGTTTTTGTTCATAGTATGCTACTTATT
CTAAAAA

>G2510 Amino Acid Sequence (conserved domain in AA coordinates:41-108)

MSPQRMKLSSPPVTINNEPTATASAVKSCGGGKETSSSTTRHPVYHGVKRRRWGKWVSEI
REPRKKSRIWLGSFPVPEMAAKAYDVAFLKGRKAQLNFPEEIEDLPRPSTCTPRDIQV
AAAKAANAVKIIKMGDDDVAGIDDDFWEGIELPELMMSGGWSPEPFVAGDDATWLVLD
GDLYQYQFMACL*

>G2893 (130..981)

AAATCATAAAAGCCTCTCTTAGTCTATTTTTATCTCACGGCTCTCTCCCCCTCTCTA
CACACACAAACACAAATAAAGCGTAAACTGAAATATTTAATTACAATTAGAAAGAGAA
CATATTAATATGTCAAATATAACAAAGAAGAAGTGTAATGGAAATGAAGAGGGTGCAGAG
CAGAGGAAAGGGCCTTGGACACTCGAGGAAGACACTTCTCTACCAATTACATTTCCCAT

TCATGATGGACATTGATTGGATAAACAATCTAATCTGATGTGTAACGTCACCTTGCACTGA
CATTTAATATGGTTTANCTATCAGTTACCTGTCTGCTTCTTGTAAGGGTATACCTTGGATC
CTTGCTCTTGAACCTGTTTTATTTAGCATGCAAA

>G39 Amino Acid Sequence (domain in AA coordinates: 24-90)
MPPSPPKSPFISSSLKGAHEDRKFKCYRGVVRKRSWGKWVSEIRVPKTRRIWLGSYDAPE
KAARAYDAALFCIRGEKGVYNFPTDKKPQLPEGSVRPLSKLDIQTATNYASSVVHVPSH
ATTLPATQVPSEVPASSDVSASTEITEMVDEYYLPTDATAESIFSVEDLQLDSFLMMDI
DWINNLII*

>G439 (128..967)

TATAAATCTTCGTTTCTACTTTTTTCTTCCATAATATAGTCAATTCGTTTTCTTAATT
AGGGCTTCTTCTCTTGTCTTCCATCTTTATTAGTTTATTTATTTTGGTTATTG
TATACAAATGGCAATGGCTTTAAACATGAATGCTTACGTAGACGAGTTCATGGAAGCTCT
TGAACCATTCATGAAGGTAACCTCATCTTCTTCTACTTCGAATTCATCAAATCCAAAACC
ATTAACCTCCTAATTCATCCCTAATAATGACCAAGTCTTACCGGTATCTAACCAAACCGG
TCCGATTGGGGCTAAACCAGCTCACTCCACACAAATCTCCAAATTCAGACAGAGTTACA
TCTCCGGCAAAACCAATCTCGTCGTCGCGCTGGTAGTCATCTTCTCACCCTAAACCAAC
CTCAATGAAGAAAATCGACGTAGCAACTAAACCGGTTAAACTATACCGAGGCGTAAAGACA
GAGGCAATGGGGTAAATGGGTAGCTGAGATTCCGGCTACCTAAAAACCGAACCCTGGTTATG
GCTCCGGTACGTTTCAAGACGGCTCAAGAAGCTGCATTAGCTTACGATCAAGCAGCTCATAA
GATCAGAGGAGACAACGCTCGTCTCAATTTCCAGACATTGTTCTGTCAGGACACTATAA
ACAGATATTGTCTCCGTCTATCAACGCAAGATCGAATCCATCTGCAATAGTTCTGATCT
TCCACTGCCTCAGATCGAGAAAACAGAACAAAACAGAGGAGGTGCTCTCTGGTTTTTCCAA
ACCGGAGAAAGAACCGGAATTTGGGGAGATATACGGATGCGGATACTCGGGCTCATCTCC
TGAGTCGGATATAACGTTGTTGGATTTCTCAAGCGACTGTGTGAAAGAAGATGAGAGTTT
CTTGATGGGTTTGCACAAGTATCCTTCTTGGAGATTGATTGGGACGCTATAGAGAACT
CTTCTGAATCCATTTTATCTTTTTTGAATTCATTTGCTCTAAATGTAGAATTTTATTTTC
AGAGCTTTGTAAGGGAAGTTCTTGAATGAGAGTTGCAGAGGACTAGTGGAACCTAACTCT
GTTTTCTTTTGTAAATGTTTATAATGGGCCGTTGAATGGGCTTATTGATTTAAACA
GCCCCAAGTTTTTAAAAA

>G439 Amino Acid Sequence (domain in AA coordinates: 110-177)
MAMALNMNAYVDEFMEALEPFMKVTSSTSSSNPKPLTPNFIPNNDQVLPVSNQTPGI
GLNQLTPTQILQIQTELHLRQNSRRRAGSHLLTAKPTSMKKIDVATKPKVLYRGVRQRQ
WGKWVAEIRLPKNRTRLWLGTFFETAQEAALAYDQAAHKIRGDNARLNFDPDIVRQGHYKQI
LSPSINAKIESICNSSDLPLPQIEKQNKTEEVLSGFSKPEKEPEFGEIYCGYSGSSPES
DITLLDFSSDCVKEDESFLMGLHKYPSLEIDWDAIEKLF*

>G470 (1..2580)

ATGGCGAGTTCGGAGGTTTCAATGAAAGGTAATCGTGGAGGAGATAACTTCTCCTCCTCT
GGTTTTAGTGACCTAAGGAGACTAGAAATGTCTCCGTCGCGCGGAGGGGCAAAAAGT
AATTCTACCCGATCCGCTGCGGCTGAGCGTGCTTTGGACCCTGAGGCTGCTCTTTACAGA
GAGCTATGGCAGCGTTGTGCTGGTCCGCTTGTGACGGTTCTTAGACAAGACGACCGAGTC
TTCTATTTTCTCAAGGACACATCGAGCAGGTGGAGGCTTCGACGAACCGGCGGCAGAA
CAACAGATGCCTCTCTATGATCTTCCGTCAAAGCTTCTCTGTGCGAGTTATTAATGTAGAT
TTAAAGGCAGAGGCAGATACAGATGAAGTTTATGCGCAGATTACTCTTCTTCTGAGGCT
AATCAAGACGAGAATGCAATTGAGAAAGAAGCGCTTCTTCTCCACCTCCGAGGTTCCAG
GTGCATTTCGTTGCAAAACCTTGACTGCATCCGACACAAGTACACATGGTGGATTTTCT
GTTCTTAGGCGACATGCGGATGAATGTCTCCACCTCTGGATATGTCTCGACAGCCTCCC
ACTCAAGAGTTAGTTGCAAGGATTTGCATGCAATGAGTGGCGATTTCAGACATATATTC
CGGGGTCAACCACGAGGCGATTTGCTACAGAGTGGGTGGAGTGTGTTTGTAGCTCCAAA
AGGCTAGTTGCAGGCGATGCGTTTATATTTCTAAGGGGCGAGAATGGAGAAATTAAGAGTT
GGTGTAAAGCGTGCGATGCGACAACAGGAAACGTGCCGCTTCTGTATATCTAGCCAT
AGCATGCATCTTGGAGTACTGGCCACCGCATGGCATGCCATTTCAACAGGGACTATGTTT
ACAGTCTACTACAAACCCAGGACGAGCCCATCTGAGTTTATTGTTCCGTTTCGATCAGTAT
ATGGAGTCTGTTAAGAATAACTACTCTATTGGCATGAGATTCAAAATGAGATTTGAAGGC
GAAGAGGCTCCTGAGCAGAGGTTTACTGGCACAATCGTTGGGATTGAAGAGTCTGATCCT
ACTAGGTGGCCAAAATCAAAGTGGAGATCCCTCAAGGTGAGATGGGATGAGACTTCTAGT
ATTCTCGACCTGATAGAGTATCTCCGTGGAAAGTAGAGCCAGCTCTTGCTCTCCTGCT
TTGAGTCTGTTCATGCTTAGGCCTAAGAGGCCAGATCAAATATAGCACCTTCATCT

CCTGACTCTTCGATGCTTACCAGAGAAGGTACAACCTAAGGCAAACATGGACCCTTTACCA
GCAAGCGGACTTTCAAGGGTCTTGCAAGGTCAAGAATACTCGACCTTGAGGACGAAACAT
ACTGAGAGTGTAGAGTGTGATGCTCCTGAGAATTCTGTTGTCTGGCAATCTTCAGCGGAT
GATGATAAGGTTGACGTGGTTTCGGGTTCTAGAAGATATGGATCTGAGAACTGGATGTCC
TCAGCCAGGCATGAACCTACTTACACAGATTTGCTCTCCGGCTTTGGGACTAACATAGAT
CCATCCCATGGTCAGCGGATACCTTTTTATGACCATTTCATCATCACCTTCTATGCCTGCA
AAGAGAATCTTGAGTGATTGAGAAGGCAAGTTCGATTATCTTGCTAACCAGTGGCAGATG
ATACACTCTGGTCTCTCCCTGAAGTTACATGAATCTCCTAAGGTACCTGCAGCAACTGAT
GCGTCTCTCCAAGGGCGATGCAATGTTAAATACAGCGAATATCCTGTCTTAATGGTCTA
TCGACTGAGAATGCTGGTGGTAACTGGCCAATACGTCCACGTGCTTTGAATTATTATGAG
GAAGTGGTCAATGCTCAAGCGCAAGCTCAGGCTAGGGAGCAAGTAACAAAACAACCCCTTC
ACGATACAAGAGGAGACAGCAAAGTCAAGAGAAGGGAAGTGCAGGCTCTTTGGCATTCTCT
CTGACCAACAACATGAATGGGACAGACTCAACCATGTCTCAGAGAAACAACCTGAATGAT
GCTGCGGGGCTTACACAGATAGCATCACCAAAGGTTTCAGGACCTTTTCAGATCAGTCAAAA
GGGTCAAAATCAACAAACGATCATCGTGAACAGGGAAGACCATTCCAGACTAATAATCCT
CATCCGAAGGATGCTCAAACGAAAACCAACTCAAGTAGGAGTTGCACAAAGGTTTACAAG
CAGGGAATTGCACCTGGCCGTTTCAGTGGATCTTTCAAAGTTCCAAACTATGAGGAGTTA
GTCGCTGAGCTGGACAGGCTGTTTGAAGTTCAATGGAGAGTTGATGGCTCCTAAGAAAGAT
TGGTTGATAGTTTACACAGATGAAGAGAATGATATGATGCTTGTGTTGGTGACGATCCTTGG
CAGGAGTTTGTGTCATGGTTTCGAAAATCTTCATATACACGAAAGAGGAAGTGAGGAAG
ATGAACCCGGGGACTTTAAGCTGTAGGAGCGAGGAAGAAGCAGTTGTTGGGGAAGGATCA
GATGCAAAGGACGCCAAGTCTGCATCAAATCCTTCATTGTCCAGCGCTGGGAACCTCTTAA
>G470 Amino Acid Sequence (domain in AA coordinates: 61-393)
MASSEVSMKGNRGGDNFSSSGFSDPKETRNVS VAGEGQKSNSTRSAAERALDPEAALYR
ELWHACAGPLVTVPRQDDRVFYFPQGHIEQVEASTNQAAEQQMPLYDLPSKLLCRVINVD
LKAEDTDEVYAQITLLPEANQDENAIKEAPLPPPPRFQVHSFCKTLTASDTSTHGGFS
VLRHHADECLPPLDMSRQPPPTQELVAKDLHANEWRFRHIFRGQPRRHLLQSGWSVVFVSSK
RLVAGDAFIFLRGENGELRVGVRRAMRQQGNVPSSVISSSHMLGLVLATAWHAISTGTMF
TVVYKPRTPSEFIVPFDQYMESVKNNYSIGMRFKMRFEGEAAPEQRFTGTIVGIEESDP
TRWPKSKWRSLSKVRWDETSSI PRPDRVSPWKVEPALAPPALSPVPMRPRKPRSNIA PSS
PDSSMLTREGTTKANMDPLPASGLSRVLQGGQEYSLRTHKTESVECDAPENSVVWQSSAD
DDKVDVVS GSRRYGSENWMSARHEPTYTDLSSGFGTNIDPSHGQRIPFYDHSSSPSMPA
KRILSDSEGKFDYLANQWQMIHSGLSLKLHESPKVPAATDASLQGRCNVYSEYPVLNGL
STENAGGNWPIRPRALNYYEEVVNAQAQAQAREQVTKQPFITQEETAKSREGNCRFLFGIP
LTNNMGTDTSTMSQRNNLNDAGLTQIASPKVQDLSDQSKGSKSTNDHREQRPFQTNPNP
HPKDAQTKTNSRSCTKVHKQGIAGRSVDLSKFQNYEELVAELDRLFEPNGELMAPKKD
WLIVYTDEENDMMLVGDDPWQEFCCMVRKIFITYTKEEVRKMNPGTLSCRSEEEAVVGEGS
DAKDAKSASNPSSLSSAGNS*

>G652 (1..606)

atgagcggaggaggagacgtgaacatgagtggtggagacagacgcaagggaaacggtgaag
tggtttgatacacagaaggggtttggtttcatcacacctagcgacggtggtgacgatctc
ttcgttcaccagtcctccatcagatctgaaggatttcgtagcctcgagctgaggaatct
gttagttcgacgttgaggttgacaactccggccgtcccaaggctattgaagtgtctgga
cccgacggtgctcccgttcagggttaacagcggtggtggtggttcattctggtggacgcggt
ggttttggcggcggtggtggaagaggagggggacgtggtggaggaagctacggaggaggt
tatggtggaagaggaagcggtggccgtggaggaggtggtggtgataattcttgctttaag
tgcggtgaaccaggtcacatggcgagagaatgctctcaagggtggtggaggatacagcgga
ggcgggggtggtggaaggtacgggtctggcggcgccggaggaggaggtggtggtggtcta
agctgctacagctgtggagagtctgggcactttgcaagggattgcactagcggtggtgct
cgttga

>G652 Amino Acid Sequence (domain in AA coordinates: 28-49, 137-151, 182-196)

MSGGGDVNMSGGDRRKGTVKWFDQKGFGITPSDGGDDL FVHQSSIRSEGFRLAAEES
VEFDVEVDNSGRPKAIEVSGPDGAPVQGNSSGGGSSGGRGGFGGGGGGRGGGGSYGGG
YGGRGSGRGGGGGDNCFKCGEPGHMARECSQGGGYSGGGGGGRYSGGGGGGGGGGL
SCYSCGESGHFARDCTSGGAR*

>G671 (61..1119)

TTCATTGAGAACAAACCCCTTTGAACTCGATCAAGAAAGCTAAGTTTGAAGAATCAAGA

ATGGTGCGGACACCGTGTTGCAAAGCCGAAGTGGGTTAAAGAAAGGAGCTTGGACTCCC
GAGGAAGATCAGAAGCTTCTCTCTTACCTTAACCGCCACGGTGAAGGTGGATGGCGAAGT
CTCCCCGAAAAAGCTGGACTCAAGAGATGCGGCAAAAGCTGCAGACTGAGATGGGCCAAT
TATCTTAGACCTGACATCAAAAGAGGAGAGTTCACTGAAGACGAAGAACGTTCAATCATC
TCTCTTACGCCCTTACGGCAACAAATGGTCTGCTATAGCTCGTGGACTACCAGGAAGA
ACCGATAACGAGATCAAGAACTACTGGAACACTCATATCAAAAAACGTTTGATCAAGAAA
GGTATTGATCCAGTTACACACAAGGGCATAACCTCCGGTACCGACAAATCAGAAAACTC
CCGGAGAAACAAAATGTTAATCTGACAACTAGTGACCATGATCTTGATAATGACAAGGCG
AAGAAGAACAACAAGAAATTTTGGATTATCATCGGCTAGTTTCTTGAACAAAGTAGCTAAT
AGGTTCCGAAAGAGAATCAATCAGAGTGTCTGTCTGAGATTATCGGAAGTGGAGGCCCA
CTTGCTTCTACTAGTACACTACTAATACTACAACCTACAAGTGTTCGGTTGACTCTGAA
TCAGTTAAGTCAACGAGTTCTTCCTTCGCACCAACCTCGAATCTTCTCTGCCATGGGACC
GTTGCAACAACACCAGTTTTCATCGAAGTTTGACGTTGATGGTAACGTTAATCTGACGTGT
TCTTCGTCCACGTTCTCTGATTCTCCGTTAACAATCCTCTAATGTACTGCGATAATTTT
GTTGGTAATAACAACGTTGATGATGAGGATACTATCGGGTTCTCCACATTTCTGAATGAT
GAAGATTTTCATGATGTTGGAGGAGTCTTGTGTTGAAAACACTGCGTTTCATGAAAGAACTT
ACGAGGTTTCTTACGAGGATGAAAACGACGTCGTTGATGTGACGCCGGTCTATGAACGT
CAAGACTTGTTTGACGAAATTGATAACTATTTTGGATGAGTGAACCTCATAATCGATGAA
TCCCACGTGACCATGTCAATATGATGTCATGATATGTTACCTTGATGATGTTGATGGT
AATAATAATAAATAATAGATGGTGATGATGACCATGCATGAATCATGAATGTAGTTCGTG
TTGTCACATATGCTTGTGTTTTTGTGTTTTTTTTTTTTTGGTCTGAAGTGTGTTGTTTCGT
TGTAATGGATTATAAATGGTGATGTAATAATTATAATGTTAAAAAATAAATAAATAAATAA
AAAA

>G671 Amino Acid Sequence (domain in AA coordinates: 15-115)
MVRTPCCKAELGLKKGAWTPEEDQKL₁₅SYLNRHGE₂₀GGWRTLPEKAGLKRCGKSCRLRWAN
YLRPDIKRGEFTEDBERSI₂₅ISLHALHGNKWSAIARGLPGRTDNEIKNYWNTHIKRLIKK
GIDPVTHKGTISGTDKSENLPKQNVNLTSDHDLNDKAKKNKNFGLSSASFLNKVAN
RFGKRINQSVLSEIIGSGGPLASTSHTTNTTTTSVSVDSSESVKSTSSSFAPTSNLLCHGT
VATTPVSSNFDVDGNVNLTCSSSTFSDSSVNNPLMYCDNFVGNNNVDDTIGFSTFLND
EDFMLEESCVENTAFMKELTRFLHEDENDVDVTPVYERQDLFDEIDNYFG*

>G779 (110..712)

GACATGCATGTAAGCATTCGGTTAATTAATCGAGTCAAAGATATATATCAGTAAATACAT
ATGTGTATATTTCTGGAAAAAGAATATATATATTGAGAAATAAGAAAAGATGAAAATGGA
AAATGGTATGTATAAAAGAAAGGAGTGTGCGACTCTTGTGTCTCGTCCAAAAGCAGATC
CAACCACAGCCCCAAAGAAGCATGATGGAGCCTCAGCCTCACCATCTCCTCATGGATTG
GAACAAAGCTAATGATCTTCTCACACAAGAACACGCAGCTTTTCTCAATGATCCTACCA
TCTCATGTTAGATCCACCTCCCGAAACCTAATTCATCTGGACGAAGACGAAGAGTACGA
TGAAGACATGGATGCGATGAAGGAGATGCAGTACATGATCGCCGTATGCAGCCCCGTAGA
CATCGACCTGCCACGTTCCCTAAGCCGAACCGCGTAACGTAAGGATAAGCGACGATCC
TCAGACGGTGGTTGCTCGTCCGCGTCGGGAAAGGATCAGCGAGAAGATCCGAATTTCTCAA
GAGGATCGTGCTGGTGGTGCGAAGATGGACACAGCTTCCATGCTCGACGAAGCCATACG
TTACACCAAGTTCTTGAAACGGCAGGTGAGGATTCTTCAGCCTCACTCTCAGATTGGAGC
TCCTATGGCTAACCCTCTTACCTTTGTTATTACCACAACCTCCCAACCTGATGAACCTAC
ACAGAAGCTCGCTAGCTAGACATTTGGTGTCTATCTCTCAACCTTT

>G779 Amino Acid Sequence (domain in AA coordinates: 126-182)
MKMENGMYKKKGVCDSVSSKSRSNHSPKRSMMEPQPHLLMDWNKANDLLTQEHAFLN
DPHHLMLDPPPETLIHLDEDEEYDEDMDAMKEMQYMIAMVQPDIDPATVPKPNRRNVRI
SDDPQTVVARRRRERISEKIRILKRIVPGGAKMDTASMLDEAIRYTKFLKRQVRILQPHS
QIGAPMANPSYLCYYHNSQP*

>G962 (148..1392)

CGTCGACTCTCTACTCAACACCACTCAATTTTCATCTCTCTTTTCCCTTCCATTGTTAGT
ATAAAAACCAAGCAAAACCTTAATCACTTTTCATCATCATATATCACCTTAATCCACATG
CATACACATATCTAGTCTTTTGGATATATGGCAATTGTATCTCCACAACAAGCATCATTT
CCCATGAGTAACCAAGTCAACAATAACGAAAAAGGTATAGAAGACAATGATCATAGAGGC
GGCCAAGAGAGTCATGTCCAAAATGAAGATGAAGCTGATGATCATGATCATGACATGGTC
ATGCCCGGATTTAGATTCCATCTTACCGAAGAAGAACTCATAGAGTTTACCTTCGCCGA
AAAGTTGAAGGCAAACGCTTTAATGTAGAAGTCACTCTTCTCGATCTTTATCGCTAT

GATCCTTGGGAACCTCTGCTATGGCGCGATAGGAGAGAAAGAGTGGTACTTCTATGTG
CCAAGAGATCGGAAATATAGAAATGGAGATAGACCGAGTAACGACTTCAGGATAT
TGGAAAGCCACCGGAGCTGATAGGATGATCAGATCGGAGACTTCTCGGCTATCGGATTA
AAGAAAACCTAGTTTCTACTCTGGTAAAGCCCCTAAAGGCACTCGTACTAGTTGGATC
ATGAACGAGTATCGTCTTCCGCACCATGAAACCGAGAAGTACCAAAAGGCTGAAATATCA
TTGTGCGGAGTGACAAAAGGCCAGGAGTAGAAGATCATCCATCGGTACCACGTTCTCTC
TCCACAAGACATCATAACCATAACTCATCGACATCATCCGTTTAGCCTTAAGACAACAA
CAACACCATTATCTCTCTCTAATCATTCCGACAACAACCTTAACAACAACAACATC
ACAACAGTAACCTGACGTTTACCATTGCTCTAGCCAATCAAAACATATATCGTCCAATG
CCTTACGACACAAGCAACAACACATTGATAGTCTCTACGAGAAATCATCAAGACGATGAT
GAAACTGCCATTGTTGACGATCTTCAAAGACTAGTTAACTACCAAATATCAGATGGAGGT
AACATCAATCACCAATACTTTCAAATTGCTCAACAGTTTTCATCATACTCAACAACAAAAT
GCTAACGCAAACGCATTACAATTGGTGGCTGCGGCGACTACAGCGACAACGCTAATGCCT
CAAACCTCAAGCGGCTTAGCTATGAACATGATTCTTGCAGGAACGATTCCAAACAATGCT
TTGTGGGATATGTGGAATCCAATAGTACCAGATGGAAACAGAGATCACTATACTAATATT
CCTTTTAAGTAATTTAATTAGATCATGATTATTATCCATGACAATAATTAATGCTGCTTT
GCGC

>G962 Amino Acid Sequence (domain in AA coordinates: 53-175)
MAIVSSTSIIPMSNQVNNEKGIENDHRGGQESHVQNEDEADDHDDHDMVMPGFRFHPT
EEELIEFYLRKVEGKRFNVELITFLDLRYDPWELPAMAAIGKEWYFYVPRDRKYRNG
DRPNRVTTSGYWKATGADRMIRSETSRPIGLKKTLLVFYSGKAPKGTRTSWIMNEYRLPHH
ETEKYQKAEISLCRVYKRPVEDHPSVPRSLSTRHHNHNSSSTSSRLALRQQQHSSSSNH
SDNNLNNNNNINLEKLSTEYSGDGSTTTTNTNSNDVTIALANQNIYRPMYPYDTSNNTL
IVSTRNHQDDDETAIVDDLQRLVNYQISDGGNINHQYFQIAQQFHHTQQQNANANALQLV
AAATTATTLMPQTGAALAMNMIPAGTI PNALWDMWNPIVPDGNRDHYTNIPFK*

>G977 (46..591)
CACCAAACCTACCTGAAACCCATTTCCATTACCATTCACACTAATGGCACGACCACAA
CAACGCTTTTCGAGGCGTTAGACAGAGGCATTGGGGCTCTTGGGTCTCCGAAATTCGTCAC
CCTCTCTTGAAAACAAGAATCTGGCTAGGGACGTTTGAGACAGCGGAGGATGCAGCAAGG
GCCTACGACGAGGCGGCTAGGCTAATGTGTGGCCCGAGAGCTCGTACTAATTTCCCATAC
AACCCTAATGCCATTCTCTACTTCTCTTCCAAGCTTCTATCAGCAACTCTTACCGCTAAA
CTCCACAAATGCTACATGGCTTCTCTTCAAATGACCAAGCAAACGCAAACACAACGCAA
ACGCAGACCGCAAGATCACAATCCGCGGACAGTGACGGTGTGACGGCTAACGAAAGTCAT
TTGAACAGAGGAGTAACGGAGACGACAGAGATCAAGTGGGAAGATGGAAATGCGAATATG
CAACAGAAATTTAGGCCATTGGAGGAAGATCATATCGAGCAAATGATTGAGGAGCTGCTT
CACTACGGTTCATTGAGCTTTGCTCTGTTTTACCAACTCAGACGCTGTGAGAAATGGCC
TTGTCGTTTTAGCGTATTCTTTTCATTTTTATTGTTTCCACAAAAACGGCGTCTGTA
GTGATGAGAGTAGTAGTGAGAGAAGGCTAATTTCAAGACATTTTGATCTGAATTGGCCTC
TTTTGAAACACTGATTCTAGTTTCTATAAGAGCAATCGATCATATGCTATGTTATGTATA
GTATTATAAAAAAATGTTATTTCTGATTNAAAAAAAAAAAAAAAAAAAAAAAAAAAAA

>G977 Amino Acid Sequence (domain in AA coordinates: 5-72)
MARPQQRFRGVRQRHWGSWVSEIRHPLKTRIWLGTFFETAEDARAYDEAARLMCGPRAR
TNFPYNPNAIPTSSSKLLSATLTAKLHKCYMASLQMTKQTQTQTQTARSQSADSDGVT
ANESHLNRGVTETTEIKWEDGNANMQNFRPLEEDHIEQMIEELLHYGSIELCSVLPTQT
L*

>G1063 (241..966)
GTAAAGAAGATGGATGGGCCACAAGTTGCTATATAAATCCTTCCACTTCTTGTGTATA
CTATTGCTTGAGTTCTGATTGGGCACAGTAGTACCATTGCCATTTCTCTCACACATACCG
TCTCTTTCTCTCATCATCAATCATCAATCATAAAGAAAAAACCTTAAATTTCACTT
GTAAGCTTTTTCACAGTTTCTCTCCATACCCATTTATCAGCTTCTCCATATCTTTCTCT
ATGGATTCTGACATAATGAACATGATGATGCATCAGATGGAGAAGCTTCTGAGTTTTGT
AACCCTAATTCCTCTTTCTCTCTCCCGACCACAACAACACTTACCCTTTTCTCTTTAAC
TCCACTCATTACCAGTCCGATCACTCAATGACCAACGAACAGGTTTCCGCTACGGTTCC
GGTTTACTCACTAACCTTCTCTCTATCTCTCCCAACACAGCTTACTCTTCCGTTTTCTT
GACAAAAGAAACAACAGTAACAACAACAATAATGGCACGAACATGGCAGCTATGCGAGAG
ATGATCTTCCGTATCGCGTGATGCAACCGATCCATATCGATCCCGAGGCGTTAAGCCA

CCGAAGAGGAGGAACGTCAGGATCTCTAAAGATCCTCAAAGCGTGGCGGCTAGGCATAGA
AGGGAGAGAATAAGCGAGAGGATTTCGGATTTTGCAACGGCTTGTTCCTGGTGGGACGAAG
ATGGATACAGCTTCGATGCTCGATGAAGCAATTCATTATGTGAAGTTTTTAAAGAAACAG
GTGCAGTCTCTGGAGGAGCAGGCGGTGGTTACTGGCGGAGGGGAGGAGGAGGAGGAAGG
GTTTTGATCGGTGGAGGTGGAATGACGGCGCGAGTGGTGGTGGTGGCGGCGGGGAGTG
GTTATGAAAAGGTGTGGAACAGTGGGGACTCATCAGATGGTGGGCAATGCACAGATTCTT
AGATGATGATGATTTTTTAATTTTATTATTATTATTAATGTTGGAGAAAAAGAGAAAA
TGATTCTGGAGAGGGAAGCCAAGTAATTTATGTGAGAGTCTTTAATTTAACTTTATTTTC
TTGTTTAGATAATGTGTAATGATGGTTTTTAAAGCCAAAGACTCTCCATGGTTGTTGGAG
CGAGTTTG

>G1063 Amino Acid Sequence (domain in aa coordinates: 131-182)

MDSDIMNMMHMQEKLPEFCNPNSFFSPDHNNTPFLFNSTHYQSDHSMNEPGRYGS
GLLTNPSSISPNTAYSSVFLDKRNNNSNNNGTNMAAMREMIFRIAVMQPIHIDPEAVKP
PKRRNVRISKDPQSVAAHRHRERISERIRILQRLVPGGTKMDTASMLDEAIHYVKFLKKQ
VQSLLEQAVVTGGGGGGGRVLIGGGGMTAASGGGGGGGVVMKGCCTVGTQMVGNAQIL
R*

>G1140 (67..729)

ATCCAAGATCCTCCAACACAGAAAGGCAGATTCAAGAACAGTAGTGAAGGAGAGATCT
GGTAAAATGGCGAGAGAGAAGATAAGGATAAAGAAGATTGATAACATAACAGCGAGACAA
GTTACTTTCTCAAAGAGAGAAGAGGAATCTTCAAGAAAGCCGATGAACCTTCAGTTCTT
TGCGATGCTGATGTTGCTCTCATCATCTTCTCTGCCACCGGAAAGCTCTTCGAGTTCTCC
AGCTCAAGAATGAGAGACATATTGGGAAGGTATAGTCTTCATGCAAGTAACATCAACAAA
TTGATGGATCCACCTTCTACTCATCTCCGGCTTGAGAAATTGTAACCTCTCCAGACTAAGT
AAGGAAGTCGAAGACAAAACCAAGCAGCTACGGAACTGAGAGGAGAGGATCTTGATGGA
TTGAACTTAGAAGAGTTGCAGCGGTGGAGAACTACTTGAATCCGGACTTAGCCGTGTG
TCTGAAAAGAAGGGCGAGTGTGTGATGAGCCAAATTTCTCACTTGAGAAACGGGGATCG
GAATTGGTGGATGAGAATAAGAGACTGAGGGATAAACTAGAGACGTTGGAAAAGGGCAAAA
CTGACGACGCTTAAAGAGGCTTTGGAGACAGAGTCGGTGACCAACAATGTGTCAAGCTAC
GACAGTGGAACCTCCCTTGAGGATGACTCCGACACTTCCCTGAAGCTTGGGCTTCCATCT
TGGGAATGAATCTGAGAGAGAGAAAGATCCAGCAGAGTTGACTTCGATGGAAGCCACAA
ATATTAAGTCTACCTTTTCCCTTTCTTTCTTTGAATAAGTGTGAAAAAGAAATTGAGAT
GGGAAGGATGAATCTCATTGCATTGCAGAGAAGCAAGTTTCAGATATTGTACGTGTTAT
TGGGTCTTTATACTATTTTTCTCCCAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA

>G1140 Amino Acid Sequence (conserved domain in AA coordinates: 2-57)

MAREKIRIKKIDNITARQVTFKRRRGIFKKADELSVLCADVALIIFSATGKLFEFSSS
RMRDILGRYSLHASNINKLMDPPSTHLRLNCNLSRLSKEVEDKTKQLRKLRGEDLDGLN
LEELQRLKLLLESLSRVSEKKGEVMSQIFSLEKRGSELVDENKRLRDKLETLEAKLT
TLKEALETESVTTNVSSYDSGTPLEDDSDTSLKLGPSWE*

>G1425 (43..1005)

ACTCTCTCAAACCATAAAAAATATTCCTCCGATCATCATTTTTAATGGAGAGTACAGATTCT
TCCGGTGGTCCTCCGCCGCGCAACCAACCTCCCTCCAGGATTCGGGTTTCATCCAACA
GACGAAGAACTTGTAATTCATTACCTCAAACGCAAAGCAGATTCTGTTCTTTACCAGTC
GCGATCATCGCCGACGTTGATCTTTACAAATTTGATCCATGGGAACTTCCCGCGAAAGCT
TCGTTTGGAACAACAAGATGGTATTTTTTTCAGTCCAAGAGATCGGAAATATCCCAACGGA
GCTAGACCTTAACCGAGCTGCGACTTCCGGTTATTGGAAAGCGACTGGTACAGATAAACCG
GTGATTTCAACCGGCGGTGGTGGTAGTAAAAAAGTGGGAGTTAAAAAGGCTCTAGTGTTT
TACAGTGGTAAACCACCAAAAGGAGTTAAATCAGATTGGATTATGCATGAATATCGGTTA
ACTGATAATAAACCTACTCACATTTGTGACTTCGGCAACAAGAAAACTCTCTCAGGCTT
GATGATTGGGTGTTGTGTCTATCTACAAGAAAAACAATAGTACAGCATCTAGACATCAT
CATCATCTTCATCATATTCATCTAGATAATGATCATCATCGTCATGATATGATGATTGAT
GATGATCGATTCCGTCATGTTCTCTGGTCTTCACTTCCCGCGGATTTTTCTGACAAT
AATGATCCGACGGCTATATATGATGGTGGCGGCGGCGGATACGGAGGTGGAAGTTACTCG
ATGAATCATGTTTTCGATCTGGATCAAAGCAGGAGCAGTTGTTTCCACCGGTGATGATG
ATGACTAGTCTAAATCAAGATTCCGGTATTGGATCGTCGTCGTCACCTAGCAAGAGATTT
AACGGCGGCGGCGGTGGAGATTGTTGACTTCTATGGCGGCGACGCCGTTAATGCAGAAC
CAAGGTGGGATTTACCAATTGCCTGGTTTGAATTGGTATTCTTGAAAACAATTTACGATG
AAGAATTTTTAAAAATTTGTGTATATATACGGTTTGAGTGATTAGGGGGCATTGGGGGA

TTTATTTACGGTTGATTATTATTGTAGTGTTATAGAACTAAGGAGATTAAATTAAATAGA
TTGGAGGAAAAAAAAAAAAAAAAAAAA

>G1425 Amino Acid Sequence (domain in AA coordinates: 20-173)
MESTDSSGGPPPPQPNLPPGFRFHTDEELVIHYLKRKADSVPLPVAIIADVDLYKFDPW
ELPAKASFGEQEWYFFSPDRKYPNGARPNRAATSGYWKATGTDKPVISTGGGGSKKVGV
KKALVFYSGKPPKGVKSDWIMHEYRLTDNKPTHICDFGNKKNLSRLDDWVLCRIYKKNNNS
TASRHHHHLHHIHLNDNHHRHDMIDDDRFRHVPPGLHFPALFSDNNDPTAIYDGGGGGY
GGGSYSNMHCFASGSKQEQLFPVMMMTSLNQDSGIGSSSSPSKRFNNGGVGDCSTSMMA
TPLMQNQGGIYQLPGLNWYS*

>G1449 (105..581)

TAGACAGAGAGAAATAGAAATAGAGAGAGAGACATGAAGAGCACTCTCAATAGAGAAG
AGAAGGAAGCATGAAGCTAGCTCTGCAGCTTCAAGGTCTCATTATGGAGGTCTCTAACT
CTTGTTCTTCATTTTCTTCATCCTCTGTGACAGTACTAAACCTTCTCCTTCTGAATCTT
CTGTTAATCTCTCCCTTAGTCTCACATTTCTTCTACTTCTCCACAAAGAGAAGCAAGAC
AAGATTGGCCACCGATAAAGTCTAGATTAAAGAGATACACTAAAGGGTCGTCGTCTTCTTC
GTCGTGGTGATGACACTTCTCTCTTTGTTAAGGTTTATATGGAAGGTGTTCCCATGGAA
GAAAACTCGACCTTTGCGTATTCTCAGGCTACGAGAGTCTATTAGAAAATCTCTCTCACA
TGTTTCGATACTTCAATCATCTGCGGTAATCGAGATCGAAAACATCATGTTTTGACATATG
AAGACAAGGATGGAGATTGGATGATGCTCGGAGATATTCCATGGGATATGTTTCTTGAAA
CCGTGAGAAGACTAAAGATCACGAGACCGGAGAGGTATTAAGAACTGGATCGGTCAAGGC
TGTGATTGCGCAGTTACGAGACGTGTAAGATTTAGGCATTGATGAAGAGACTTGAGGCGG
GACGGAGCTATTGCTGCATATTGCAACAAAGGCCCTTGAAGAAGTTGGAGAATTGATTGAT
GCATATATTTATTTATATGACACCTTTGAGTGTGTTTTTCTTATAAATAAATCACAATA
TCCAAGACTTCTCTTTAAA

>G1449 Amino Acid Sequence (domain in AA coordinates: 48-53,74-107,122-152)

MEVSNSSCSFSSSSVDSTKPSPESSVNLSSLTFPSTSPQREARQDWPPPIKSRLRDTLK
MRLLLRGDDTSLFVKVMEGVPIGRKLDLCVFSGYESLLENLSHMFDTSIICGNRDRKH
HVLITYEDKGDWMMVGDIPWDMFLETVRRLKITRPERY*

>G1897 (1..678)

ATGCCTTCTGAATTCAGTGAATCTCGTCGGGTTCTTAAGATTCCCCACGGCCAAGGAGGA
TCTGTTGCGATTCCGACGGATCAACAAGAGCAGCTTTCTTGTCCTCGCTGTGAATCAACC
AACACCAAGTTCTGTTACTACAACAACCTACAACCTCTCACAACCTCGTCATTTCTGCAAG
TCTTGTCGCCGTTACTGGACTCATGGAGGTACTCTCCGTGACATTTCCCGTCGGTGGTGT
TCCCGTAAAAGCTCAAAACGTTCCCGGACTTATTCCTCTGCCGCTACCACTCCGTTGTC
GGAAGCCGGAACTTTCCCTTACAAGCTACGCCTGTTCTTTTCCCTCAGTCGTCTTCCAAC
GGCGGTATCACGACGGCGAAGGGAAGTGCTTCGTGCTTCTATGGCGGTTTCAGCTCTTTG
ATCAACTACAACGCCCGCTGAGCAGAAATGGGCCTGGTGGCGGGTTTAAATGGGCCAGAT
GCTTTTGGTCTTGGGCTTGGTCACGGGTCGTATTATGAGGACGTCAGATATGGGCAAGGA
ATAACGGTCTGGCCGTTTCAAGTGGCGCTACTGATGCTGCAACTACTACAAGCCACATT
GCTCAAATACCCGCCACGTGGCAGTTTGAAGGTCAAGAGAGCAAAGTCGGGTTCTGTGTCT
GGAGACTACGTAGCGTGA

>G1897 Amino Acid Sequence (domain in AA coordinates:34-62)

MPSEFSESRRVPKIPHGQGSVAIPTDQEQLSQPRCESTNTKFCYNNYNFSQPRHFCK
SCRRYWTHGGTLRDIPVGGVSRKSSKRSRTYSSAATTSVVGSRNFPLQATPVLFPQSSN
GGITTAKGSASSFYGGFSSLINYNAAVSRNGPGGGFNGPDADFGLGLGHGSYYEDVRYGQG
ITVWPFSSGATDAATTTSHIAQIPATWQFEGQESKVGFSVSGDYVA*

>G2143 (89..784)

TCTTCTTCTTCTCCATACCTTATCTCACCAGCTTCTCCATATCTCTCAAAGAAAAACA
AACCCTATAAATTCCACAAAAAGGAGGATGGATAACTCCGACATTCTAATGAACATGAT
GATGCAGCAGATGGAGAAGCTTCTGAACACTTCTTAACCTCAAACCTAACCCTAATCC
CCATAACATTATGATGCTTTCTGAATCCAACACCCACCCGTTCTTCTTCAACCCCACTCA
TTCTCATCTCCCATTGACCAAACCATGCCTCACCACCAACCCGGTTAAATTTCCGGTA
CGCCCCCTCCCCGTATCATCTCTCCCGGAGAAGAGAGGAGGCTGCAGCGACAACGCCAA
CATGGCGCGGATGAGAGAGATGATCTTTCGAATAGCCGTGATGCAGCCTATACATATTGA
TCCGGAATCCGTAAAGCCACCAAGAGAAAGAACGTGAGGATCTCTAAGGATCCACAGAG
CGTGGCAGCTCGGCATCGAAGGGAGAGGATAAGCGAGCGGATTCCGATTCTTCAGCGGCT
TGTTCCCGGTGGGACTAAGATGGATACGGCGTCGATGCTCGATGAGGCTATCCATTACGT

TAAGTTTCTCAAGAAGCAAGTGCAGTCGCTGGAGGAACATGCGGTGGTTAACGGCGGAGG
AATGACGGCGGTGGCCGGAGGAGCACTTGCGGGTACTGTTGGTGGAGGATATGGAGGAAA
AGGGTGTGGCATTATGCGGTCTGATCATCACCAGATGCTTGGAAATGCACAGATTCTTAG
ATGATGATGATGTTGATTTTTAAATATATATCATATGTTTATTAATATGACGGGAAAAAA
TATTATCGAGGGAGTTGAATTTAGTATCATGAACTATGAGAGCATTTTTTTTAAATGTT
TTTATCTTTCCGGGTTTCGATAATGTTTGGGATGGTTAATTAACAATTTAAAAGTCAGAC
AACTTGGTTGTAAGACTAAAGAATAAGCATAGTTTATCAATTTATCATTACTAAATGAA
ATAG

>G2143 Amino Acid Sequence (domain in aa coordinates: 128-179)

MDNSDILMNMMMQOMEKLEPHFSNSNPNNPNPHNIMMLSESNTHPFFNPTHSHLPFDQTM
PHHQPLNFRYAPSPSSSLPEKRGCSNANMAAMREMIFRIAVMQPIHIDPESVKPPKR
KNVRISKDPQSVAAHRHRERISERIRILQRLVPGGTKMDTASMLDEAIHYVKFLKKQVQS
LEEHAVVNGGMTAVAGGALAGTVGGGYGGKGCIMRSDHHQMLGNAQILR*

>G2535 (1..1005)

ATGAACATATCAGTAAACGGACAGTCACAAGTACCTCCTGGCTTTAGGTTTCACCCAACC
GAGGAAGAGCTCTTGAAGTATTACCTCCGCAAGAAAATCTCTAACATCAAGATCGATCTC
GATGTTATTCCTGACATTGATCTCAACAAGCTCGAGCCTTGGGATATTCAAGAGATGTGT
AAGATTGGAACGACGCCGCAAAACGATTGGTACTTTTATAGCCATAAGGACAAGAAGTAT
CCACCCGGGACTAGAACCAACAGAGCCACCACGGTCGGATTTTGGAAAGCGACGGGACGT
GACAAGACCATATATACCAATGGTGATAGAATCGGGATGCGAAAGACGCTTGTCTTCTAC
AAAGGTCGAGCCCCCTCATGGTCAGAAATCCGATTGGATCATGCACGAATATAGACTCGAC
GAGAGTGTTAATCTCCTCGTGTGGCGATCATGACGTCACGTCAGAAACGTGTGATGTC
ATAGGAAGTGACGAAGGATGGGTGGTGTGTCGTGTTTTCAAGAAAAATAACCTTTGCAAA
AACATGATTAGTAGTAGCCCGGCGAGTTTCGGTGAACCGCGTTCGTTCAATGAGGAGACT
ATCGAGCAACTTCTCGAAGTTATGGGGCAATCTTGTAAGGAGAGATAGTTTACAGCCCT
TTCTTAAACCTCCCTAACCTCGAATGCCATAACAACACCACCATCACGAGTTATCAGTGG
TTAATCGACGACCAAGTCAACAACCTGCCACGTCAGCAAAGTTATGGATCCCAGCTTCATC
ACTAGCTGGGCCGCTTTGGATCGGCTCGTTCCTCACAGTTAAATGGGCCCCAACTCGTAT
TCAATACCAGCCGTTAATGAGACTTCACAATCACCGTATCATGGACTGAACCGGTCCGGT
TGTAATACCGGTTTAAACACCAGATTACTATATACCGGAGATTGATTTATGGAACGAGGCA
GATTTTCGCGAGAACGACATGCCACTTGTGTAACGGTAGTGGATAA

>G2535 Amino Acid Sequence (conserved domain in AA coordinates:11-114)

MNISVNGQSQVPPGFRFHPTEEBLLKYLRKKISNIKIDLDVIPDIDLNKLEPWDIQEMC
KIGTTPQNDWYFYSHKDKKYPTGTRTNRATTGFWKATGRDKTIYTNDRIGMRKTLVVFY
KGRAPHGQKSDWIMHEYRLDESVLISCGDHDVNVETCDVIGSDEGWVVCRVFKNNLCK
NMISSSPASSVKTPSFNEETIEQLLEVMGQSKGEIVLDPFLKLPNLECHNNTTITSYQW
LIDDQVNNCHVSKVMDPSFITSWAALDRLVASQLNGPNSYSIPAVNETSQSPYHGLNRS
CNTGLTPDYIPEIDLWNEADFARTTCHLLNGSG*

>G2557 (94..1215)

TCGACTTCTGTGAACATCTGTTTGTCTCTTCTCCGGTTTCACCTTTTTCATGTCCT
GCCGTTATTACAACGAGGATTGTGTTGATCCGATGGAAGGATTGGAATCTGTGTACGCT
CAAGCTATGTATGGAATGACACGAGAGAGCAAAATCATGGAGCATCAAGGATCAGATTG
ATTTGGGGAGGAAATGAGCTAATGGCTCGAGAACTCTGTTCTTCTTCTTATCACCAC
CAACTCATTAAATCCGAATCTTAGCAGCTGTTTCATGTCTGATCTTGGAGTCTTAGGTGAG
ATTCAACAGCAGCAACATGTTGGCAACAGAGCTAGCTCGATAGATCCATCATCACTCGAT
TGTTTGTATTCTGCGACGTCGAATAGCAACAACACCTCGACGGAGGACGATGAAGGAATA
TCTGTGCTTTTCTCAGATTGTCAGACTCTTTGGAGCTTTGGTGGAGTCTCATCTGCAGAG
TCTGAGAACAGAGAGATCACTACTGAGACGACAACAACGATAAAGCCTAAGCCTTTGAAG
AGAAACAGAGGAGGAGATGGAGGAACTACTGAGACTACAACAACAACAAAACCTAAG
TCTTTGAAGAGAAAACAGAGGAGACGAGACAGGAAGTCACTTTAGTCTTGTTCATCTCAA
GATGATTCCGAGAAAGGAGGTTTCAAGCTTATATACGATGAGAATCAATCGAAATCAAAG
AAACCAAGAACAGAGAAAGAACGAGCGGTTCTTCAACATTAGTTTCCAACATTCAACT
TGTTTGTCTGACAAATGTCGAGCCCGATGCTGAGGCGATTGCACAAATGAAGGAGATGATA
TACAGAGCGGCTGCATTTAGACCGGTGAATTTCCGGTTAGAGATTGTGGAGAAGCCTAAG
AGGAAGAACGTCAGATATCGACGGATCCTCAAACGGTTTCAGCGAGACAGAGAAGGGAG
AGGATAAGTGAGAAAGATTAGGGTTTACAAAACATTGGTTCCAGGTGGGACGAAGATGGAT
ACTGCATCAATGCTTGATGAAGCTGCTAATTATCTCAAGTTTCCTTAGAGCACAAAGTAAAA

GCTTTAGAAAACCTTGAGACCCAAGCTTGACCAACCAATCTCTCTTTCTCTCTGCTCCT
ACATCGTTTCCATTATTCCACCCATCTTTTCTTCCATTGCAAAATCCTAATCAAATCCAT
CATCCAGAGTGTGACAGATTATAAACTTTTGAGTTTCATCATCATCAACAGAATCATGG
CGTCTTGATTGTTTTAGCAGTTCTCAAGAAAGGCAACTTCTGTGACAAGGGTGGTGTCCG
GCAGTGTGTTTACACTTTCCAGTCTTTGTTTTGCATTTCTTTTTATATAAAGTTGTAT
TTTATATAGAATCTGTGGAATTCGAGGGTTGAAATATTGTGAAAAACAGAGCCGCAAGAG
GTTAATTACAGTCTCTGCAATATTTCAACCTTTTATTACTTTATTAGAGTAAAGATAGC
GT

>G2557 Amino Acid Sequence (domain in aa coordinates: 278-328)
MEGLSEVYAQAMYGMTRESKIMEHQGSDLIWGGNELMARELCSSSSYHHQLINPNLSSCF
MSDLGVLGEIQQQHVGNRASSIDPSSLDCLLSATSNSNNTSTEDDEGISVLFSDCQTLW
SFGGVSSAESENREITTTETTTTIKPKPLKRNRRGGDGGTTETTTTTTKPKSLKRNRRGDET
SHFSLVHPQDDSEKGGFKLIYDENQSKSKPRTEKERGGSSNISFQHSTCLSDNVEPDAE
AIAQMKEMIYRAAAFRPVNFGLEIVEKPKRKNVKISTDPQTVARQRRERISEKIRVLQT
LVPGGTKMDTASMLDEAANYLKFLRAQVKALENLRPKLDQTNLSFSSAPTSFPLFHPSFL
PLQNPNQIHHEC*

>G259 (52..786)
GAGATCTTCTACTACTTGTGTTTTCTTCAAGAATAATAATTTTCGTTTTATATATGGAAGAT
GCTGGTGAACATTTACGGGTGTAACGATAACGTTAACGACGAGGAGCGTTTGCCATTGGAG
TTTATGATCGGAACTCAACATCCACGGCGGAGCTACAGCCGCTCCACCGTTCTTGGA
AAGACATACAAAGTGGTGGAGGATCCGACGACGACGGGTTATATCTTGGAAACGAATAC
GGAAGTGGTTTTCGTCGTGTGGCAGCCGGCAGAATTCGCTAGAGATCTGTTACCAACACTT
TTCAAGCATTGCAACTTCTCTAGCTTCGTTCCGACGCTCAATACTTACGGTTTTCGAAAA
GTAACGACGATAAGATGGGAATTTAGTAATGAGATGTTTCGAAAGGGGCAAAGAGAGCTT
ATGAGCAATATCCGAAGAAGGAAGAGCCAACATTGGTCACACAACAAGTCTAATCACCAG
GTTGTACCAACAACAACGATGGTGAATCAAGAAGGTCATCAACGGATTGGGATTGATCAT
CACCATGAGGATCAACAGTCTTCCGCCACTTCATCCTCTTTCGTATACACTGCATTACTC
GACGAAAAACAAATGCTTGAAGAATGAAAAAGAGTTATTAAGCTGCGAACTTGGGAAAAAC
AAGAAGAAATGCAAGCAGCTTATGGAGTTGGTGGAGAGATACAGAGGAGAAGACGAAGAT
GCAACTGATGAAAGTGATGATGAAGAAGATGAAGGGCTTAAGTTGTTCCGAGTAAACTT
GAATGAACTAGATTGCTAGATTGATATTCGTAATATACAGTTTCTTCATATTCTTAGA
AGTTTTGCATACTATATATAGTACTCTTTTAAGACATGCAAGATCAGAACATATG

>G259 Amino Acid Sequence (domain in AA coordinates: 27-131)
MEDAGEHLRCDNVNDEERLPLEFMIGNSTSTAELOPPPPFLVKTYKVVEDPTDGVISW
NEYGTGFVVWQPAEFDLLPTLFKHCFNFSSFVRQLNTYGFVKVTTIRWEFSNEMFRKGQ
RELMSNIRRRKQSHWSHNKSNHQVVPVTTMVNQEGHQIRIGIDHHHEDQSSATSSSFVYT
ALLDENKCLKNENELLSCELGKTKKKCKQLMELVERYRGEDEDATDESDDDEDEGLKLFG
VKLE*

>G353 (82..570)
ACCAAACCTCAAAAACACAAACCACAAGAGGATCATTTTCATTTTTATTGTTTCGTTTTA
ATCATCATCATCAGAAGAAAAATGGTTGCGATATCGGAGATCAAGTCGACGGTGGATGTC
ACGGCGGCGAATTGTTGATGCTTTTATCTAGAGTTGGACAAGAAAACGTTGACGGTGGC
GATCAAAAACGCGTTTTCACATGTAAACGTTGTTGAAGCAGTTTCATTTCGTTCCAAGCC
TTAGGAGGTACCGTGCGAGTCACAAGAAGCCTAACAACGACGCTTTGTGCTCTGGATTG
ATGAAGAAGGTGAAAACGTCGTCGCATCCTTGTCCCATATGTGGAGTGGAGTTTCCGATG
GGACAAGCTTTGGGAGGACACATGAGGAGACACAGGAACGAGAGTGGGGCTGCTGGTGGC
GCGTTGGTTACACGCGCTTTGTTGCCGAGCCCCACGGTGACTACGTTGAAGAAATCTAGC
AGTGGGAAGAGAGTGGCTTGTGTTGGATCTGAGTCTAGGGATGGTGGACAATTTGAATCTC
AAGTTGGAGCTTGGAGAAGAGTTTATTGATTTTATTTATTTTCTTAAATTTTCTGAAT
ATATTTGTTTCTCTCATCTTTGAATTTTCTTAAATTTCTAGATTATACATACATCCGC
AGATTTAGGAACTTTCATAGAGTGAATCTTTCTTTCTGTAAAAATATATTTTACTTG
TAGCAA

>G353 Amino Acid Sequence (domain in aa coordinates: 41-61, 84-104)
MVAISEIKSTVDVTAANCLMLLSRVGQENVDDGDQKRVFTCKTCLKQFHSFQALGGHRAS
HKKPNNDALSSGLMKVKTSSHPICGVVEFPMQALGGHMRRHRNESGAAGGALVTRAL
LPEPTVTTLLKSSSGKRVACLDLSLGMVDNLNLKLELGRVTVY*

>G354 (27..533)

CCTAGAAGTCACTAAGTCGATTCAAAATGGTTGCGAGAAGTGAGGAAATGTGATAGTGG
AAGAAGATACGACTGCGAAATGTTTGATGTTGTTATCAAGAGTCGGAGAATGCGGCGGCG
GCTGCGGGGAGATGAACGTGTTTTCCGATGCAAGACTTGTCTTAAAGAGTTCTCATCGT
TTCAAGCTTTGGGAGGTTCATCGTGCAAGCCACAAGAACTTATCAACAGTGACAATCCAT
CACTTCTTGGATCCTTGTCCAACAAGAAAATAAAACGTCTCATCCTTGTCCGATATGTG
GAGTGAAGTTTCCGATGGGACAAGCTCTTGGTGGTCACATGAGGAGACATAGGAACGAGA
AAGTCTCAGGCTCGTTGGTTACACGTCTTTTCTACCGGAGACGACGACGGTGACGGCTT
TGAAGAAATTTAGTAGTGGGAAGAGAGTGGCTTGTGTTGGATTGGACTTAGATTCGATGG
AGAGTTTGGTCAATTGGAAGTTGGAGTTGGGAAGAACGATTTCTTGGAGTTAAGTTTTTG
GGTTGTATACAGTTTCACATGATTTTGTAACTTTTGTGATCCAATTATCGTACCGATCG
ATGTGAATATTATTTTGATACAATAAAA

>G354 Amino Acid Sequence (domain in aa coordinates: 42-62, 88-109)

MVARSEEIVIVEEDTTAKCLMLLSRVGECGGGCGGDERVFRCKTCLKEFSSFQALGGHRA
SHKKLINSNDNPSLLGSLSNKTKTSHPCPICGVKFPMGQALGGHMRHRNEKVSGLSVTR
SFLPETTTVTALKKFSSGKRVACLDDLDSMESLVNWKLELGRITISWS*

>G638 (86..1861)

GAATTAAGGTTTAAACCTTTACCTTTTTTCCCTTCACTATCGATAATTGATCTTCTCT
TTCCGGCTGAATATAAATCTGAAAAATGGATCAAGATCAGCATCCTCAGTACGGTATACC
GGAGCTCCGGCAGCTCATGAAAGCGGAGGAAGGACGACTACTACAACACCGTCTACTTC
TTCTCATTTTTCCCTCTGATTTCTTCGGTTTTTAACTTGTCTCCGGTGCAGCCACCGCCACA
CCGTCTTCATCAGTTCACTACTGATCAAGATATGGGTTTCTTGCCACGTGGCATAACATGG
ATTGGGTGGAGGTTCTTCAACGGCTGGAATAACAGTAACTTAAACGCGAGTACTAGTGG
TGGAGGAGTTGGGTTTAGTGGGTTTCTTGACGGTGGTGGTTTCCGCGAGCGGAGTAGGAGG
AGACGGTGGAGGAACCTGGAAGGTGGCCGAGACAAGAAACCCTAACCTCTGTTGGAATTAG
ATCTCGTCTTGATCATAAATTCAAAGAAGCTAATCATAAAGGACCTCTTTGGGATGAAGT
TTCTAGGATTATGTCCGAGGAACATGGATACCAAAGGAGTGGGAAGAAATGCAGAGAGAA
GTTTGAGAATCTGTACAAATACTATAGTAAGACTAAAGAAGGCGAAGCCGGAAGACAAGA
CGGAAAACATCACAGATTTTTCCGCCAGCTCCAAGCGCTATACGGGGATTCTAATAACTT
GGTTTCTTGTCCCAATCATAAACACGCAGTTTCATGAGCAGTGCTCTTCATGGTTTCCATAC
TCAAAACCTATGAACGTTGCTACAACAACGTCCAACATCCATAACGTTGATAGTGTTCAT
TGGTTTTTCATCAAAGCCTTAGTCTTTCTAACAACCTACAACCTCCTCCGAGCTTGAGCTGAT
GACTTCTCTTTCGGAAGGGAAATGATTCTAGTAGTAGAAGGAAAAAGAGGAGTTGGAAAGC
GAAGATAAAGGAGTTCATTGATACGAACATGAAAAGGTTGATAGAGAGGCAAGATGTTTG
GCTTGAGAAGTTGACAAAGGTTATTGAAGACAAAGAGGAACAACGGATGATGAAAGAAGA
GGAATGGAGGAAGATTGAAGCTGCAAGGATTGATAAAGAGCATTGTTTTGGGCTAAAGA
GAGGCGAGGATGGAAGCTAGGGATGTTGCGGTGATTGAGGCATTGCAATACTTGACAGG
AAAGCCATTGATAAAGCCGCTGTGTTTCATCCCCGGAAGAGAGGACAAATGGTAATAATGA
GATCCGAAAACATAGTGAGACACAGAATGAGAATGGAAGCGATCAAACGATGACTAACAA
TGTTTTGTGTTAAAGGAAGTAGTAGCTGCTGGGGTGAGCAAGAGATTTTAAAGCTTATGGA
GATAAGAACGAGCATGGACTCGACCTTTCAAGAGATATTAGGAGGGTGCTCGGATGAGTT
TCTATGGGAGGAAATCGCAGCGAAGTTGATTAGTTAGGGTTTGATCAGAGAAGTGCCTT
ATTATGCAAGGAAAAGTGGGAATGGATAAGCAATGGAATGAGGAAAGAAAAGCAAGCAAT
CAACAAGAAAAGAAAGGATAATTCTGTCAGCTGCGGCGTGACTACCCGAGAAACGAAGA
AAATCCAATCTACAATAATCGAGAAAGTGGATATAATGATAATGATCCGCATCAAATCAA
CGAACAAGGCAATGTAGGTTCTTCAACATCAAACGCAACGCAACGCAACGTAACCAC
TGGAAATCCGAGCGGTGCAATGGCTGCTAGTACAACTGCTTCCCGTTCTTCATGGGAGA
TGGAGATCAGAAATTTGTGGGAGAGTTATGGTTTGAGGCTCAGTAAAGAAGAGAATCAGTA
AGTAATTTCTCTTAATGAAGAAGAAGGTAATCATGTGGTTAACTAATTTCTTTGAGT
TAGCTATATATGAGATAAACCTTGACTTAGCTATTATATGTCACATGCTGCTTAGAATTA
AGAAATATTTGTGGGGCTTAACGAATTATATATCAGCATATATAAGATGAGAGTCTAAG
AATTATATCAAAATTAGGCTTTAACCACGTACGATTATATATTATGTTTTCATGTATTTA
TTCTGTAAGACTTTTTTAATATCAATCTTTCTCTAAA

>G638 Amino Acid Sequence (domain in AA coordinates: 119-206)

MDQDQHPQYGIPELRQLMKGGGRTTTTTTPSTSSHFPSPDFGFNLAPVQPPPHRLHQFTTD
QDMGFLPRGIHGLGGSSSTAGNNSNLNASTSGGGVGFSGFLDGGGFGSGVGGDGGGTGRW
PRQETLTLLEIRSLDHKFKEANHKGPLWDEVSRIMSEEHGYQSRGKKCREKFENLYKYY
SKTKEGEAGRQDGKHRFFRQLQALYGDSSNNLVSCPNNHTQFMSSALHGFHTQNPMMNVAT

TTSNIHNVDSVHGPHQSLSLNNYNSSSELELMTSSSEGNDSRRKRKRSWKAKIKEFIDT
NMKRLIERQDVWLEKLTKVIEDKEEQRMMEEWKIEAARIDKEHLFWAKERARMEARD
VAVIEALQYLTKGLIKPLCSSPEERTNGNNEIRNNSETQNGSDQTMNNVCVKGSSS
CWGEQEILKLMBERTSMDSTFQELGGCSDEFLWEBIAAKLIQLGFDQRSALLCKEKWEW
ISNGMRKEKKQINKRKNDNSSCGVYYPNEENPIYNNRESGYNDNDPHQINEQNGVSS
TSNANANANVTGNPSGAMAASTNCFPFMGDGDQNLWESYGLRLSKEENQ*

>G869 (428..1402)

AGGAACAGTGAAAGGTTTCGGTTTTTTGGGTTTCGATCTGATAATCAACAAGAAAAAGGG
TTTGATTTATGTCCGCTGGGTTTGAATCGACTGTGATTTGTCTTTGATTCATATCTCTT
CTCCGATTTTCATCATCATCTTCCCATCATCGTCGTCTTTGAAATCTTGCTTCTCAACG
CTCTTCACTTCTGCTGTAATAAGCAGAGGCTTGTCTGGAGACTCCTTCTCTTCCATGC
GCTTAAGACCCAAAAGGACTTGTCTAGTGTTGAAGTCTTTGGGGTTTTACATAAAGC
AGCAAAAGTTTTCTTTTTTCATAGTTCGCTGAGAGTTTTGAGTTTTGATACCAAAAAGT
TTTGACCTTTTTAGAGTGATTTTTTGTCTTTCTGTTTTCTGGGTATTTTTGAGGAGTGGG
TTTAACAATGGTTGCGATTAGAAAGGAACAGTCTTTGAGTGGTGTAGTAGCGAGATTAA
GAAGAGAGCTAAGAGAAACACTCTATCGTCCCTTCCCTCAAGAAACCAACCTTTGAGGAA
AGTCCGTATTATTGTGAATGATCTTATGCTACTGATGATTCCTCTAGTGATGAGGAAGA
GCTTAAGGTTCCTAAGCCAAGGAAATGAAACGTATCGTTCGTGAGATTAACTTTCTCTT
TATGGAAGTTTCTGAACAGCCTTCTGAGAGTTCTTCTCAGGACAGTACTAAACTGATGG
CAAGATAGCTGTGTGAGCTTCTCTGCTGTTCTTAGGAAGAAGCCTGTTGGTGTAGGCA
AAGGAAATGGGGGAAATGGGCTGCTGAGATTAGAGATCCTATTAAGAAAACCTAGGACTTG
GTTGGGTACTTTTTGATACCTTTGAAGAAGCTGCTAAAGCTTATGATGCTAAGAAGCTTGA
GTTTGATGCTATTGTTGCTGGAAATGTGTCCACTACTAAACGTGATGTTTCTTCATCTGA
GACTAGCCAATGCTCTCGTTCCTCACCTGTTGTTCTGTTGAGCAAGATGACACTTCTGC
ATCAGCTCTCATTGTGTCAACAACCTGATGACGCTCTGACCGTTGCTCCAACCTGCTCC
AACTCCAAATGTTCTGCTGGTGGAAACAAGGAAACGTTGTTGATTTGACTTTACTAA
TCTACAGATCCCTGATTTTGGTTTCTTGGCAGAGGAGCAACAAGACCTAGACTTCGATTG
TTTCTCGCGGATGATCAGTTTGATGATTTTCGGCTTGCTTGATGACATTCAAGGATTGCA
AGATAACGGTCCAAGTGCCTTACCAGATTTCGACTTTGCGGATGTTGAAGATCTTCAGCT
AGCTGACTCTAGTTTCGGTTTTCTTGTATCAACTTGCTCCTATCAACATCTCTTGCCCAT
AAAAAGTTTTGCAGCTTCATAGGATCTTGCTTAGTAATGTTAAGTGAGAAGAGTGTTTTG
TTTTTTCGTTTATGCTTTAGTAATTAAGACATACAAAAGTGTGTGTTCCGGATTGTAGT
AAGATCTTAAGACATAAAGCCGGTTTTGCAATTAGGAATCGAGTTTTAATGAAGTTTGA
GTTTATGTTTG

>G869 Amino Acid Sequence (domain in AA coordinates: 109-177)
MVAIRKEQSLSGVSSEIKKRAKRNLTSSLPQETQPLRKVRIIVNDPYATDDSSSDEBELK
VPKPRKMKRIVREINFPSEVSEQPSESSSQDSTKTDGKIAVSASPAVPRKPKVGVQRK
WGKWAABIRDPIKKTRTWLGTFTDLEEAAYDAKKLEFDAIVAGNVSTTKRDVSSSETS
QCSRSSPVVPVEQDDTSASALTCVNNPDDVSTVAPTAPTNPVAGGNKETLFDFTNLQ
IPDFGFLAEQQLDFDCFLADDQFDDFGLLDDIQGFEDNGPSALPDFDFADVEDLQLAD
SSFGLDQLAPINISCLPKSFAAS*

>G1645 (25..1104)

CGTCGACCTCCCAACACTAATCCATGTTTATAACGGAAAAACAAGTGTGGATGGATGAG
ATCGTCGCAAGAAGAGCTTCTTCTTCTTGGGACTTCCCTTTCAACGACATTAATATTCAT
CAGCATCATCATCGTCACTGCAACACAAGTCATGAGTTTGAAATCTTGAAGAGTCCTCTT
GGAGATGTAGCGGTTACGAAGAAGAGAGTAATAATAATAACCTAATTTCAAGTAACAGC
GAGAGTGGTAAGAAGGAGACAACAGATAGTGGTCAGTCTTGGTCCTCGTCGTCTTCAAAA
CCATCGGTCTTGGGGAGAGGACATTGGAGACCAGCTGAAGATGTTAAACTCAAAGAGCTT
GTCTCCATTTACGGCCCAAAAACCTGAACCTCATAGCTGAAAAGCTTCAAGGAAGATCT
GGGAAGAGCTGTAGACTACGATGGTTTAAACCAATTGGACCCGAGGATAAACCGAAGAGCT
TTCACAGAAGAAGAAGAGGAGAGGCTGATGCAAGCACATAGGCTTTATGGTAACAAATGG
GCAATGATTGCGAGGCTTTTCCCTGGTAGAACTGATAATTCACTGAAGAACCATTGGCAT
GTTGTTCATGGCTCGTAAGTATAGAGAACACTCTTCTGCTTACCGTAGGAGAAAGCTTATG
AGTAATAATCCACTTAAACCTCACCTACCAATAATCATCATCTAACCCTAACCCTAAT
TACCACTCTTTTATCTCCACTAATCATTACTTCGCTCAGCCTTTCCCGAGTTTAAATTG
ACTCATCACCTGGTTAATAATGCCCTATCACGAGTGACCATAACCAGCTTGTGTGCTCT
TTCCATTGCTTTCAAGTTATGAGAACAATGAACCTCCGATGGTTGTGAGTATGTTTGGC

AACCAAATGATGGTTCGGCGATAACGTTGGTGCCACGTGACAGCGGTTATGCAATATTCCG
CACATTGACCCCTAGTAACCAAGAGAAACCGGAGCCAAATGATGCAATGCATTGGATCGGA
ATGGACGCGGTAGATGAGGAGGTGTTGAAAAGGCTAAGCAGCAACCACATTTTTTCGAT
TTTCTTGGCTTGGGGACGGCGTGAATGTTGAACAAATTGGTGTTAATCAGATAACGACAG
TGGC

>G1645 Amino Acid Sequence (domain in AA coordinates: 90-210)
MFITEKQVWMDIVARRASSSWDFPFNDINIHQHHHRHCNTSHEFEILKSP LGDVAHVHEE
ESNNNNPNFNSSESKKETDSDGQSWSSSSSKPSVLGRGHWRAEDVKLKLVS IYGPQN
WNLIAEKLQGRSGKSCRLRFNQLDPRINRRAFTEEBEERLMQAHRLYGNKWAMIARLFP
GRTDNSVKNHWHVVMARKYREHSSAYRRRKLMSNNPLKPHLTNNHHPNPNPNYHSFISTN
HYFAQPFPEFNLTHHLVNNAPITSDHNQLVLPFHCFOGYENNEPPMVVSMFGNQMMVGDN
VGATSDALCNIPHDPSNQEKPEPNDAWHWIGMDA VDEEVFEKAKQOPHFDFLGLGTA*

>G1038 (240..1574)
GCTCGTTTTCAAATTA AAAACAGGGAGAAATTTGGAAATTCAGTACGACGGGAGATAAA
ACCTAACATACGCCATGGTGACCGTTATCTAAACTACGCCAAAATATTTGAAGTGTCTGTC
GTTTTATAATAAAACGCAAAACAAAACCCACTCCCACTTTCTCCTTTCCAAAAAAGAAC
TCTCGCCACTTTCTCTGCTCTTTCTTCTCTCTCTCTTTCTTGTTCGCCGCGCATCA
TGGAGAAAAGCGGCTTCTCTCCCGTGGTCTAAGGGTTCTTGTCTGATACGATGATCCAA
CTTGGCTCAAGATTTCTCGAGAAAATGCTCAAGAAGTGTCTTACGAAGTAACGACCTGTG
GATTAGCTAGAGAGGCTTTGAGGTTGCTGAGGGAGCGTAAAGATGGATATGATATCGTGA
TCAGCGATGTGAACATGCCTGACATGGATGGTTTCAAGCTTCTTGAGCATGTTGGTCTTG
AATTAGACCTCCCTGTAATAATGATGTCTGGTGGACGGCGAAACAGCCGAGTGTGAAGG
GAGTGCACACGGGAGCTTGTGATTACCTCTTGAAGCCGATAAGAATGAAGGAGTTAAAGA
TTATATGGCAACATGTTCTGAGAAAGAGCTTCAAGAAGTGAGAGATATCGAAGGCTGTG
GATACGAAGGAGGAGCGGATTGGATCACTCGATACGATGAAGCACATTTTCTTGGAGGTG
GTGAAGATGTTTCTTTTGGGAAAAAGAGAAAAAGACTTTGACTTTGAGAAGAAGCTTCTTC
AAGATGAGAGTGTATCCATCATCTTCTTCTTCCAAGAAAGCTAGAGTTGTTTGGTCTTTTG
AGCTTCATCATAAGTTTGTCAACGCCGTTAACCAAATCGGATGCGATCACAAAGCTGGTC
CCAAGAAGATATTGGATCTCATGAATGTTCCATGGCTCACTAGAGAAAATGTTGCAAGCC
ACCTTCAGAAATATAGACTTTACCTGAGCAGATTAGAGAAAGGAAAGGAGCTCAAGTGT
ATTCAGGTGGCGTGAAGAAATGCGGATTCATCTCCAAAAGATGTGCAAGTGAATTCAGGCT
ACCAAAGCCCTGGGAGGAGCAGCTATGTATTCTCTGGAGGAAATCTCTGATCCAAAAG
CAACAGAGATTGATCCAAAGCCACTTGCTTCAGCTTCTTTGTCTGACCCCAACACCGATG
TGATCATGCCTCCGAAAACAAAAGACGCGTATAGGATTTGATCCTCCCATTTCTCTCCT
CTGCGTTTGACTCTCTGCTTCTTGGAAATGATGTTCCAGAGGTCCTTGAATCGAAGCCGG
TTCTGTATGAGAATAGCTTTCTCCAGCAACAACCATTGCCAAGTCAAAGTTCTTATGTTG
CAATTTCTGCACCATCTCTCATGGAGGAGGAAATGAAGCCTCCTTATGAGACACCAGCAG
GAGGCAGTAGTGTAATGCAGATGAGTTTCTCATGCCACAAGACAAGATCCCTACTGTAA
CCCTTCAAGATTTGGATCCCTCTGCCATGAAGCTGCAGGAGTTCAACACAGAAGGCGATT
CTGAAGAAGCTTGAAGTGGGGAACCTCCAGAATCACATCATTCTGTTTCTTTAGACACTG
ACTTAGACTTGACTTGGCTTCAAGGCGAGCGTTTCTTGCAAAACACCGACTCCAGTTTCAA
GATACAGTAGTAGCCCATCACTCTATCTGAGCTCCAGCCACCTTAATTGGTATGGAA
ATGAGCGGCTGCCTGACCCTGACGAGTATTCCTTCATGGTAGACCAAGGTTTATTATAT
CTTAACCTTGTTCATAAATCTTTCTGATATTTGGTGGTGTATGCAGAAAGATTTT
GTGGGTATACCTGAAAATAATCTTGCTTTCCCAAGAACCTTCCATGATCGGATGCATTGT
ACAATAATCCACGAGTGTCTGATGCTAATTACACCAAACAGGTTGATGACAGTGATAAGG
CCACATGTTTCACACCGCTCGCTTAAGATCTTTACTGTACCTGGAAGGAAA

>G1038 Amino Acid Sequence (domain in AA coordinates: 198-247)
MEKSGFSPVGLRVLVDDPTWLKILEKMLKKCSYEVTTCLAREALRLRLREKDG YDIV
ISDVNMPDMDGFKLEHVHLELDLPVIMMSVDGETSRVMKGVHTGACDYLLKP IRMKELK
IIWQHVLRRKKLQEVRIEGCGYEGADWITRYDEAHFLGGGEDVSFGKKRKDFEFKLL
QDESDPSSSSSKKARVVWSFELHHKFNVAVNQIGCDHKAGPKKILDLNVPWL TRENVAS
HLQKYRLYLSRLEKGLKCYSGGVKNADSSPKDVEVNSGYQSPGRSSYVFSGNSLIQK
ATEIDPKPLASASLSDPNTDVIMPCKTKTRIGFDPPISSSAFDSLPLWNDVPEVLESKP
VLYENSFLQQQLPLSQSSYVAISAPSLMEEEMKPPYETPAGGSSVNADEFLMPQDKIPTV
TLQDLDP SAMKLQEFNTEGDSEEA*

>G1073 (62..874)

[illegible]

ccaaaggcaagaggttagagcttctgctggcaatattacctgataacaacggttcacttt
atgggtgatcttaagagaatctgtgaaaccgagcttggtttgatattctcaatgttgctca
caaaacatgtgttcaagattagcaaacagtagtcttagtagacgccataagctgtagaattccactgg
ttagcgatataaccgagaatcatttttggcgagacgtgactcacccagagaacggggaag
agtcaagcccttcaatcgctgctgtgtgtgcttctcaagactggcctgaagtgacaaaat
atgcggttttagtttgctcaagctcacaggcaagaactatacaagatttgtataaaa
catggcaagatcctgttcgcggtactgttagtggcggtatgatcaggacaccttctatct
catttagaaaaagcaacagggcaaaaaccgcttcgaattatctttatcgtagtgagtaa
gccaagggcaattctatcaagttttactctatgagttggatgcaattcgaaaggcttggtg
catcgcttgaaccgaattatcagccaccggtgacattcatagttgtacagaagcgtaacc
acactcgtttgtttgctaataatcaccgagacaaaaacagtagtaccgaagcggaata
tcttaccaggtactgttagttagactaaaatatgtcatccaactgaattcgacttctacc
tttgtagccatgcgggtattcagggacaagcaggcctgcacattaccatgttctttggg
acgagaacaatttcacagcagatggtattcaatctctgactaacaatctctgttatacct
atgcgcggtgcactcggtcggtctctatagttcctccagcgtattatgctcatcttgag
catttcgagcacgtttctacctggaacctgagataatgcaagacaacggatcaccgggta
aaaagaacacgaaaacaacaactgtcggagacgtaggtgtgaagcctttaccagccttga
aggagaatgtgaagagagtaattgttctactgctaaaaatccaaacattccttaatcagtt
ttaataagtagtttggttgtttgcttagtgcggcttagatttaccatgtttttctt
atgtaaattttgcggttgggtttagcctttaggaattagtgattagggtttttctaa
agttgtacttttagctgatgataacggttgatgcagtgactttgttaaaacctctcttcta
cagtagtgtttacgtcggtcctc

>G1146 Amino Acid Sequence (domain in AA coordinates: 886-896)

MPIRQMKDSSETHLVIKTQPLKHHNPKTVQNGKIPPPSPSPVTVTPATVTQSQASSPSP
PSKNRSRRNRGRKSDQGDVCMRPSRRPKPPPSQTTSSAVSVATAGEIVAVNHQMOM
GVRKNSNFAPRPGFTLGTKCIVKANHFLADLPTKDLNQYDVTITPEVSSKSVNRRAIAE
LVRLYKESDLGRRPLPAYDGRKSLYTAGELPFTWKEFSVKIVDEDDGIINGPKRERSYKVA
IKFVARANMHHLGEFLAGKRADCPQEAQVILDLVRELSVKRFPVGRSFFSPDIKTQR
LGEGLSWCGFYQSIRPTQMGLSLNIDMASAAFIEPLPVIEFVAQLLGKDVLSKPLSDSD
RVKIKKGLRGVKVEVTHRANVRKRYRAGLTTPQTRLMFPVDENCTMKSIVBYFQEMYG
FTIQHTHLPCLQVGNQKKASYLPMEACKIVEGQRYTKRLNEKQITALLKVTQRAEGQRN
DILRTVQHNAVDQDPYAKEFGMNISEKLASVEARILPAPWLKYHENGKEKDCLPQVGQWN
MMNKKMINGMTVSRWACVNFSSRVQENVARGFCNELGQMCEVSGMEFNPEPVIPIYSARP
DQVEKALKHVYHTSMNKTGKLELELLAILPDNNGSLYGLDKRICETELGLISQCCLTKH
VFKISKQYLADVSLKINVKMGGRNTVLVDAISCRIPLVSDIPTIIFGADVTHPENGEES
PSIAAVVASQDWPEVTKYAGLVCAQAHQELIQDLYKTQDPVRGTVSGGMIRDLLISFR
KATGQKPLRIIFYRDGVSEGFYQVLLYELDAIRKACASLEPNYQPPVTFIVVQKRHHTR
LFANNHRDKNSTDRSGNILPGTVVDTKICHPTEFDLYLCSHAGIQGTSRPAHYHVLWDEN
NFTADGIQSLTNLCYTYARCTRSVSIVPPAYYAHLAAFRRARFYLEPEIMQDNGSPGKKK
TKTTTTVGDVGVKPLPALKENVKRVMFYC*

>G1267 (152..967)

AAGTAGAGAATAATAATCACATCAAGATTGTTTATAACCTCCCCNTAATCACCTTCTTA
NTNACCACCTCTCCGGCTCTCAACAGAACAAACAAAAAACAGCTCCGTTGTCCTG
TTCCGGCGAAATCGGACGGTCGAGATCAATCATGCATCGTAGAGCAGCAATTCAAGAATC
GGATGACGAAGAAGATGAGACTTACAACGACGTCGTTCTTGAATCTCCTTCTTCTGTGA
AGACTCAAAGATCTCAAACCAACTCAAAGAAAAGGAGGAACGTAGAGAAGAGAGTTGT
CTCAGTTCCGATAGCTGACGTGGAAGGATCTAAGAGCAGAGCGGAAGTATATCCACCGTC
CGATTTCATGGGCTTGGAGAAAGTACGACAAAAACCGATCAAAGGCTCGCCTTATCCAG
GGGATATTACAGATGTAGTAGCTCAAAGGATGTCCGGCGAGGAAGCAGGTGGAGAGAAG
CCGTGTGGACCTTCTAAGCTTATGATTACTTACGCCTGCGACCACAATCACCTTTCCC
TTCTCTCTCCGTAACACCAAATCCCACCGCTCCTCCGTCGTCCTCAAACCGCAAA
GAAAGAGGAAGAATACGAAGAGGAGGAAGAACTAACCCTCACCGCCGAGAGGAACC
ACCGGCGGGACTTGATCTAAGCCACGTAGACTACCGTTGCTATTAGGCGGCTGCTACAG
CGAAATCGGAGAGTTCGGGTGGTTCTACGACGCGTCGATCTCATCATCTGTTCTTC
GAATTTCTCGACGTAATCTAGAGAGAGGTTTTTCAGTAGGCCAAGAGGAAGATGAGTC
TTTGTTCGGTGATCTCGGTGATTTACCTGATTGCGCCTCCGTGTTCCGCGGTGGGACTGT

TGCGACGGAGGAGCAACATCGAAGATGTGATTTTGGCGCCATTCTTTCTGTGATAGTTC
TAGATGAGTTTGTGTGTGTAGCCAAAACCAAGAAAAAACACAATTTTTTTATTTTCC
ACTGTAAAGGTGTATCAATGGTGGATTCATTTTTTAAAAA

>G1267 Amino Acid Sequence (domain in AA coordinates: 70-127)
MHRRAAIQESDDEEDETYNDVVPESPSSCEDSKISKPTPKRRNVEKRVVSVPIADVEGS
KSRGEVYPPSDSWAWRKYGQKPIKGSPPYRGYYRCSSSKGCPARKQVERSRVDPSKLMIT
YACDHNHPFPSSSANTKSHRSSHVLKTAKKEEYEEEEELTVTAAEEPPAGLDLSHVD
SPLLLGGCYSEIGFGEWFDASISSSSGSSNFLDVTLEGRFSVGQEEDESLFGDLGDLPD
CASVFRRGTVATEEQHRRCDFGAIPFCDSSR*

>G1269 (88..951)

AACAATTCTCTCTCTTTATTCTTCTTCTTCAGCTTCAGATTTTCAGATCTTAAATCTTC
AAGTCTTCTTCTTCTTCTTCTGCAACCATGGCTATGCAGGAACGTTGTGAGAGTTTATGT
TCTGATGAACCTTATATCTTCTCAGATGCCTTTTACCTCAAGACAAGAAAGCCTTATACC
ATCACTAAACAAAGAGAGAAATGGACAGAAGCAGAGCATGAGAAGTTTGTAGAAGCATTG
AAACTCTATGGCAGAGCTTGGAGACGAATCGAAGAACATGTTGGAACAAAACTGCAGTT
CAGATTCGAAGCCATGCGCAGAAAGTTCTTTACTAAGGTTGCTCGCGATTTTGGTGTAGC
TCTGAGTCCATTGAGATCCCGCTCCAAGGCCAAAGAGAAAGCCGATGCATCCTTACCCCT
AGAAAGCTTGTGATTCTCTGATGCAAAAGAGATGGTATACGCTGAACCTAACCGATCCAAG
CTGATTGAGGATGAAGATAACCGATCTCAACATCGGTTTATCAGCTCATGGCTCAGAT
GGATTAGGTTCCATTGGTTCAAATTCACCTAATCTTCTTCAGCTGAGTTATCATCTCAC
ACAGAGGAATCATTTGCTCTAGAAGCAGAGACCAACAGAGCCTTAAGCTCTTTGGA
ACTTTTGTAGTTGGTGTATCAACTCTTCAATGAGTTGTGATGATTCTGAAGATGGCAAG
AAGAAGCTATACTCAGAAACACAGTCTCTTCAATGTTCTTCTTCTACTTCAGAAACGCT
GAAACAGAAGTGGTAGTGTGCGGAGTTCAAAGAAGTGAGAGATCAGCTTTCTCTCAGTTA
AAATCGTGGTGACTGAGATGAACAACATGAGAGGGTTCATGCCTTACAAAAGAGAGTA
AAGGTGGAAGAAACATTGACAATGTAAATATCATATCCTTTGTGGTGAAGTGTTCGT
TTGTGTCAAGTCAGTTGTGTAAGTCTTTTGGATTAGATTGTAAGATCACTCCAAAGTTTCGTGT
CAGAGTATTAGGGAAGTTTGTGGATTAGATTGTAAGATCACTCCAAAGTTTCGTGT
CTTTCCATATAACAGTTAGAAATTGAGATCCTTGTACTTAAACATTTTTATTGTATCAA
TCAAATCTTCTGATGAAAAA

>G1269 Amino Acid Sequence (domain in AA coordinates: 27-83)
MAMQERCESLCSDELISDDAFYLKTRKPYTITKQREKWTEAEHEKFVEALKLYGRWRR
IEEHVGTKTAVQIRSHAQFFTKVARDFGVSSSEIEIPPPRPRKPKMHPYPRKLVIPDAK
EMVYAEITGSKLIQDENDRSPSVLSAHGSDGLGSIGNSPNSSSABLSHTEESLSLEA
ETKQSLKLFKFTFVVGDDYNSMSDCDDSEDGKKLYSETQSLQCSSSTSENAETEVEVSEF
KRSESAFSQLKSSVTEMMNMRGFMPYKKRVKVEENIDNVKLSYPLW*

>G1452 (175..1296)

ATTTATTAAGCATCAATGAGAGAACTTCAGAGCTGGGTTTGAGTTCTGTCCAATAATACA
TAACCACGTTATCATTTTTGTCTTTACTATCTCATTACACTCTTCTGTTATTTCGCCCCA
TTCTTACAGTCATTACTCTCTATAGGGCTCGAGCGGCCCGCGGAGGTTTCTATGCAG
ATGGTTACACTTCCCGCTCCATTGCCCAGATTGGGTTCGGTGTAAAGTCGCAATTAGTA
CTCACTATAGGGCTCGAGCGGCCCGCGGAGGTAAGATCAAACAATGTCTAAAGAA
GCTGAGATGTCTGATCGCGGTGTCGGCTTTGTTCCCTGGTTTTAGATTCTCTCTACTGAT
GTTGAACTTATCTGCTACTATCTTCGTCGTAAATTCGATGGTGTGAGAACTCTGTTGCT
GTGATTGCTGAGGTGAGATTTACAAGTTCGAGCCGTGGGACTTGCCAGAGGAATCGAAA
CTGAAATCGGAGAACGAGTGGTTTTACTTCTGCGCGAGGGGAGGAAGTACCCGCACGGG
TCACAAAGCCGCGAGCCACACAGCTAGGATATTGGAAGCGACCGGTAAAGAGCGGAGT
GTTAAATCCGGGAACCAAGTTGTTGGAACCAAGAGAACGCTTGTATTTTCATATCGGTTCG
GCTCCTCGTGGCGAGAGAACGGAGTGGATTATGCATGAATACTGCATCCATGGAGCCCCA
CAGGATGCATTAGTGGTGTGCGGTTAAGAAAAATGCTGATTTTCGGGCTAGTTCGACC
CAAAAAATTGAGGATGGTGTGCAAGACGATGGCTACGTTGGCCAAAGAGGTGGTTTG
GACAAGGAGGACAAATCTACTATGAATCTGAGCATCAGATACCAATGGTGACATCGCA
GAATCATCAATGTTGTTGAGGATCAGGCCGATACCGATGATGATTGTTACGCCGAGATT
CTGAACGATGATATAATAAGCTCGACGAAGAAGCGTTGAAAGCTAGCCAAGCGTTTCGA
CCAATAATCCAATCATCAAGAAACAATATCAAGCGAGTCATCGAGTAAGAGGTCAAAA
TGTGGTATAAAAAAAGAATCAACGGAAACAATGAATTGTTACGCTTTGTTTCAGGATCAAG
AACGTTGCCGGAACCGACTCCAGCTGGAGATTCCCGAACCCGTTCAAAATCAAGAAAGAT

GATAGCCAGAGATTGATGAAGAATGTTCTGGCCACTACTGTTTTCTTGGCTATCTTATTT
TCTTTCTTTTGGACTGTATTAATAGCTAGGAACATAAGCTAGTTACGACATACATATTAT
TTATACATAAATAATATAGTATTTTGTCTATGGCAAAAAAAAAAAAAAAAAA

>G1452 Amino Acid Sequence (domain in AA coordinates: 30-177)
MQMVHTSRISIAQIGFVKSQLVLTIGLERPPGQVKDQMSKEAEMSIIVSALFPGFRFSP
TDVELISYYLRRKIDGDNESVAVIAEVEIYKFEPWDLPEESKLKSENEWFYFCARGRKYP
HGSQSRRTQLGYWKATGKERSVKSQNVVGTKRTL VFHIGRAPRGERTEWIMHEYCIHG
APQDALVVCRLRNADFRASSTQKIEDGVVQDDGVYVQGRGLDKEDKSYESEHQIPNGD
IAESSNVVEDQADTDDDCYAEILNDDI IKLDEEALKASQAFRPTNPTHQETISSESSSKR
SKCGIKKESTETMNCYALFRIKNVAGTDSSWRFPNPFKIKKDDSQLRMKNVLATTVFLAI
LFSFFWTVLIARN*

>G1494 (114..1406)
TCGACAGAGTTGTGTGGGCGTGGAACCTTGGACTAGTTCACATATCAGGTTATATAGAT
CTTCTCTTTCAACTTCTGATTCGTCAGAAAGCTTTCTTAATCTGAGATCTGACATGGAAC
ACCAAGGTTGGAGTTTGTAGGAGAATTATAGTTTGTCCACTAATAGAAGATCTATCAGGC
CACAAGATGAAGTGTGGAGTTATTATGGCGAGATGGACAAGTGGTCTGCGAGAGCCAAA
CTCATAGAGAACAAACCCAAACCCAGAAACAAGATCATCATGAAGAAGCCCTAAGATCCA
GCACCTTTCTTGAAGATCAAGAACTGTCTCTTGGATCCAATACCCCTCCAGATGAAGACC
CATTCGAACCCGACGACTTCTCTCCCACTTCTTCTCAACCATGGATCCCTCCAGAGAC
CAACCTCAGAGACGGTTAAGCCTAAGTCCAGTCTCTGAACCTCCTCAAGTCATGGTTAAGC
CTAAGGCCTGTCTCTGACCTCCTCTCAAGTCATGCCTCCTCCAAAATTTAGGTTAACAA
ATTTCATCATCGGGGATTAGGGAAACAGAAATGGAACAGTACTCGGTAACGACCGTTGGAC
CTAGCCATTGCGGAAGCAACCCATCACAGAAGCATCTCGATGTCTCAATGAGTCATGATC
GAAGCAAAAACATAGAAAGAAAGCTTAATCCGAACGCAAGTTCCTCATCAGGTGGCTCCT
CTGGTTGCAGCTTTGGCAAAGATATCAAAGAAATGGCTAGTGGAGATGCATCACAAACCG
ACCGTAAGAGAAAACGTATAAATCACACTGACGAATCTGTATCTCTATCAGATGCAATCG
GTAACAAGTCGAACCAACGATCAGGATCAAACCGAAGGAGTCGAGCAGCTGAAGTTCATA
ATCTCTCCGAAAGGAGGAGGAGAGATAGGATCAATGAGAGAATGAAGGCTTTGCAAGAAC
TAATACCTCAGTCAGTAAACTGATAAAGCTTCGATTTTAGACGAAGCCATAGATTATT
TGAAATCACTTCAGTTACAGCTTCAAGTGATGTGGATGGGGAGTGAATGGCGGCGCGG
CGGCTTCGGCTCCGATGATGTTCCTCGGAGTTCAACCTCAGCAGTTCATACGTCAGATAC
AGAGCCCGGTACAGTTACCTCGATTTCCGGTTATGGATCAGTCTGCAATTCAGAACATC
CCGTTTGTAGTTTGCAAAACCCCGGTACAAAACAGATCATCTCCGACCGGTTTGTAGAT
ACATCGGTGGGTTCCACACATGCAGGCCGCGACTCAGATGCAGCCGATGGAGATGTTGA
GATTTAGTTACACGGCGGGACAGCAAAGTCAACAACCGTCGTCTGTGCGGACGAAGACCA
CCGACGGTTCTCGTTTGGACCACTAGGTTGGTGAGCCACTTTGC

>G1494 Amino Acid Sequence (domain in aa coordinates: 261-311)
MEHQGSFEENYSLSTNRRSIRPQDELVELLWRDQVVLQSQTHREQTQTQKQDHHEEAL
RSSTFLEDQETVSWIQYPPDEDPFEPDDFSSHFFSTMDPLQRPTSETVKPKSSPEPPQVM
VKPKACPDPPPQVMPPPKFRLTNSSSGIRETEMEQSVTTVGP SHCGSNPSQNDL DVSM
HDRSKNIEKLNPNASSSSGGSSGCSFGKDIKEMASGRCTTDRKRKRINHTEDES VLSLSD
AIGNKSNQRSGSNRRSRAAEVHNLSERRRRDRINERMKALQELIPHCSKTDKASILDEAI
DYLKSLQLQLQVMMWMSGMAAAAASAPMMFPGVQPQQFIRQIQSPVQLPRFPVMDQSAIQ
NNPGLVCQNPVQNQIISDRFARYIGGFPHMQAATQMOPMEMLRFSSPAGQQSQQPSSVPT
KTTDGSRLDH*

>G1548 (1..2511)
ATGGCAATGTCTTGCAAGGATGGTAAGTTGGGATGTTTGGATAATGGGAAGTATGTGAGG
TATACACCTGAACAAGTTGAAGCACTTGAGAGGCTTTATCATGACTGTCTTAAACCGAGT
TCTATTCCGCCGTCAGCAGTTGATCAGAGAGTGTCTTATCTCTTAACATTGAGCCTAAA
CAGATCAAAGTGTGGTTTTCAGAACCGAAGATGTAGAGAGAAACAAAGGAAAGAGGCTTCA
CGGCTTCAAGCTGTGAATCGGAAGTTGACGGCAATGAACAAGCTCTTGATGGAGGAGAAT
GACAGGTTGCAGAAGCAAGTGTACAGCTGGTCCATGAAAAAGCTACTTCCGTCAACAT
ACTCCAAATCCTTCACTCCCAGCTAAAGACACAAGCTGTGAATCGGTGGTGACGAGTGGT
CAGCACCAATTGGCATCTCAAATCCTCAGAGAGATGCTAGTCTTGCAGGACTTTTGTCC
ATTGCAGAAGAACTTTAGCAGAGTTTCTTCAAAGGCACTGGAACCGCTGTTGAGTGG
GTTTCAGATGCCTGGAATGAAGCCTGGTCCGATTCATTTGGAATCATCGCTATTTCTCAT
GGTTGCACTGGTGTGGCAGCACGCGCCTGTGGCCTAGTGGGTCTTGAGCCTACAAGGGTT

GCAGAGATTGTCAAGGATCGTCTTCGTGGTTCCGCGAATGTCGAGCTGTTGAAGTTATG
AACGTGTTGCCAACTGCCAATGGTGAACCGTTGAGCTGCTTTATATGCAGCTCTATGCA
CCAACCTACATTGGCCCCACCACGCGATTTCTGGCTGTTACGTTACACCTCTGTTTTAGAA
GATGGCAGCCTTGTGGTGTGCGAGAGATCTCTAAGAGCACTCAAATGGTCCTAGTATG
CCACTGGTTTCAAGATTTTGTGAGAGCAGAGATGCTTTCCAGTGGGTACTTGATACGGCCT
TGTGATGGTGGTGGCTCAATCATACACATAGTGGATCATATGGATTTGGAGGCTTGTAGC
GTGCCCTGAGGTCTTGCGCCCGCTCTATGAGTCACCCAAAGTACTTGCACAGAAGACAACA
ATGGCGGCACTGCGTCAGCTCAAGCAAATAGCTCAGGAGGTTACTCAGACTAATAGTAGT
GTTAATGGGTGGGGACGGCGTCTGCTGCCCTTAAGAGCTCTCAGCCAGAGGCTAAGCAGA
GGCTTCAATGAAGCTGTAAATGGTTTCACTGATGAAGGATGGTCAGTGATAGGAGATAGC
ATGGATGATGTCACAATCACTGTAAACTCTTCTCCAGACAAGCTAATGGGTCTAAATCTT
ACATTTGCCAATGGCTTTGCTCCTGTAAGCAATGTTGTTTTATGCGCAAAAGCATCAATG
CTTTTACAGAATGTTCTCCGGCGATCCTGCTTCGGTTTCTGAGGGAGCATAGGTCAGAA
TGGGCTGACAACAACATTGATGCGTATCTAGCAGCAGCAGTTAAAGTAGGGCCTTGTAGT
GCCCCAGTTGGAGGATTTGGAGGGCAGGTTATACTTCCACTTGCTCATACTATTGAGCAT
GAAGAGTTTATGGAAGTCATCAAATTGGAAGGTCCTTGGTCATTCCCTGAAGATGCAATC
GTTCCAAGAGATATCTTCTTCTTCAACTTTGTAGCGAATGGATGAAAATGCTGTAGGA
ACCTGTGCGGAACCTTATATTTGCTCCAATCGATGCTTCGTTTGGCGGATGATGCACCTCTG
CTTCTCTTCTGTTTTCGTATTATCCCTCTTGATTCCGCAAGGAAGTATCTAGCCCAAAC
CGAACCTTGGATCTTGCTTCCGCACTGGAAATGGTTCAGCTGGAACAAAAGCCTCAACT
GATCAATCAGGAACTCCACATGTGCAAGATCTGTGATGACAATAGCATTGAGTTTGGT
ATCGAGAGCCATATGCAAGAACATGTAGCATCCATGGCTAGGCAGTATGTTGAGGTATC
ATATCATCGGTGCAGAGAGTAGCATTGGCTCTTCTCCTTCTCATATCAGCTCACAAAGT
GGTCTACGCACTCCTTTGGGTACTCCTGAAGCCCAAACACTTGCTCGTTGGATTGCCAG
AGTTACAGGGGCTACATGGGTGTTGAGCTACTTAAATCAAACAGTGACGGCAATGAATCT
ATTCCTTAAGAATCTTTGGCATCAGCTGATGCTATAATCTGCTGCTCAATGAAGGCCTTG
CCCGTCTTACATTTGCAAAACAGGCGGCACTTGACATGCTGGAGACTACATTAGTTGCT
CTTCAAGACATCTCTTTAGAGAAGATATTTGATGACAATGGAAGAAAGACTCTTTGCTCT
GAGTTCCACAGATCATGCAACAGGGCTTCGCGTGCCTTCAAGGCGGGATATGTCTCTCA
AGCATGGGGAGACCAGTTTTCGTATGAGAGAGCAGTTGCTTGGAAAGTACTCAATGAAGAA
GAAATGCTCATTGCATCTGCTTTGTGTTTCATCAATTGGTCTTTGTGTGA

>G1548 Amino Acid Sequence (domain in AA coordinates: 17-77)
MAMSKDGLKGLDNGKYVRYTPEQVEALERLYHDCPKPSSIRRQQLIRECPILSNIEPK
QIKVWFQNRRCREKQKREASRLQAVNRKLTAMNKLMEENDRLQKQVSQLVHENSIFYRQH
TPNPSPKADTSCSVVTSQGHQLASQNPQRDASPAGLLSIAETLAEFLSKATGTAVEW
VQMPGMPKPGPDSIGIIAISHGCTGVAARACGLVGLPTVRAEIVKDRPSWFRECRAVEVM
NVLPTANGGTVELLYMQLYAPTTLAPPRDFWLLRYTSVLEDGSLVVCERSLKSTQNGPSM
PLVQNFVRAEMLSSGYLIRPCDGGGSIHIVDHMDLEACSVPEVLRPLYESPKVLAQKTT
MAALRQLKQIAQEVQTQNSSVNGWGRPAALRALSQLSRGFNEAVNGFTDEGWSVIGDS
MDDVTITVNSSPDKLMGLNLTTFANGFAPVSNVVLCAKASMLLQNVPPAILLRFLREHRSE
WADNNIDAYLAAAVKVGPC SARVGGFGGQVILPLAHTIEHEEFMEVIKLEGLGHSPEDAI
VPRDIFLLQLCSGMDENAVGTCAELIFAPIDASFADDAPLLPSGFRIIPLDSAKEYSSPN
RTLDLASALEIGSAGTKASTDQSGNSTCARSVMTIAFEFGIESHMQEHVASMARQYVRGI
ISSVQRVALALSPSHISSQVGLRTPPLGTPEAQTARWICQSYRGYMGVELLKSNSDGNES
ILKNLWHHTDAIICCSMKALPVFTFANQAGLDMLETTLVALQDISLEKIFDDNGRKTLC
BFPQIMQQGFACLQGGICLSSMGRPVSYERAVAWKVLNEEENAHCICFVFINWSFV*

>G1574 (1..1962)

ATGGATGATACAATGGACATGAGTTCAGGTAGTGATGAAGAAGTACAAGAAGAGAAGACC
ACTGTTAACGAGAGGGTCATCTATCAGGCTGCATTACAAGATCTGAAGCAACCCAAGACC
GAAAAGGATCTACCTCCTGGTGTCTTACAGTTCCTCTTATGAGGCATCAGAAAATTGCA
TTGAACTGGATGCGTAAGAAAGAAAAAGAACGAGGCACTGTTTGGGAGGGATATTAGCA
GATGATCAGGGACTTGGTAAACGATCTCGACGATCTCTCTTATCCTGTTACAAAAGTTG
AAGTCACAATCAAAGCAGAGAAAGCGAAAAGGTCAAACCTCTGGTGGTACATTGATTGTT
TGTCAGCAAGTGTGTAACAATGGGCAAGAGAAGTTAAAGAGAAGGTTTCTGATGAA
CACAACTCTCTGTTTTAGTCCACCATGGATCTCACAGAACCAGATCCAACAGAAATA
GCAATATATGATGTGGTCATGCAACTTACGCCATTGTTACAAATGAAGTTCCACAAAAC
CCTATGCTGAATCGTTATGATAGTATGAGAGGCAGAGAAAGCCTTGACGGATCGAGTTTG

ATTTCAGCCTCACGTTGGTGCCTAGGAAGAGTTAGGTGGTTGAGAGTAGTATTAGATGAA
GCTCATACAATTAATAAACCATAGAACCTTAATTGCAAAAGCTTGTTTTAGCCTTAGAGCC
AAAAGGAGATGGTGTGTTGACTGGAACGCCGATAAAGAACAAGTAGACGATCTTTATAGC
TATTTCAGATTTCTTAGATATCATCCATATGCCATGTGCAATTCATTTACCAAAGAATC
AAAGCTCCAATTGATAAAAAGCCTCTTCATGGTTACAAGAAGCTTCAAGCTATTCTAAGG
GGTATAATGTTGCGCCGCACCAAAGAATGGTCTTTCTACAGGAAGCTTGAATTGAATTCA
CGTTGGAAGTTTGAGGAATATGCTGCTGATGGGACTTTGCATGAACACATGGCTTATCTT
TTGGTGATGCTTTTGCCTACGCAAGCTTGTAACCATCCACAACCTTGTTAACGGATAT
AGTCACTCAGATACTACAAGAAAATGTCAGATGGAGTTCGAGTAGCCCCCTAGAGAGAAT
CTAATCATGTTCTCGATCTCTTGAATTAATCCTCAACCACCTGCTCTGTTTGTAGTGAT
CCACCAAAGACCTGTTGTTACTTTGTGTGGCCATGTGTTTTGTTATGAGTGTGTGTCT
GTAAACATTAACGGGGATAACAATACGTGCCCTGCACCTTAATTGCCACAGCCAGCTTAAA
CATGATGTTGTTTTCTAGTATCTGCAGTTAGAAGTTGCATCAACGATTATGATGATCCT
GAAGATAAAAATGCTTTAGTTGCATCAAGGCGAGTTTATTTTCATCGAAAATCCGAGCTGT
GATAGAGATTCTTCAGTCGCTTGAGAGCAAGGCAGTCCAGACACTCCACCAATAAAGAC
AATAGTATCAGTGGACTGAATCTCATTTTTACGTTTCTCAAAGACAAATGTAATGATTAT
GAAACAGGTGCGATGTTGATGTCTCTTAAAGCTGGAACCTTGGAATTGAATATGGTAGCT
GCAAGTCATGTCATTCTACTGGACCTATGGTGAATCCAACAACAGAGGATCAAGCTATT
GATCGAGCTCATCGTATCGGACAAACTCGAGCTGTTACGGTCACTCGTATTGCCATCAAA
AATACCGTTGAGGAACGAATTTTGACTCTTCATGAACGTAAGGAACATTGTTGCATCT
GCATTGGGTGAAAAAACTGGCAAAAGTTCTGCGATTCAACTAAGACTAGAAGATCTCGA
ATATCTGTTTTTTGGTGTGTAGAATATCCAGAGTTTTTATTGATAAGAGGAATAAACC
TTTAGCTATTTAATAAGTCACAAGTGTGAATGTAATGAATAA

>G1574 Amino Acid Sequence (domain in AA coordinates: 28-350)
MDDTMDMSSGSDEEVQEEKTTVNERVIYQALQDLKQPKTEKDLPPGVLTVPMLRHQKIA
LNMWRKKEKRSRHLGGILADDQGLGKTIISTISLILLQKLKSQSKQRKRKGQNSGGTLIV
CPASVVKQWAREVKEKVSDEHKLSVLVHHGSHRTKDPTEIAIYDVVMTTYAIVTNEVPQN
PMLNRYDSMRGRESLDGSSLIQPHVGALGRVRLRVVLDEHTIKNHRTLIAKACFSLRA
KRRWCLTGTPIKNKVDDLYSYFRFLRYHPYAMCNSFHQRIKAPIDKKPLHGYKKLQAILR
GIMLRRTKEWSFYRKLELNSRWKFEEYAADGTLHEHMAYLLVMLLRRLRQACNHPQLVNGY
SHSDTTRKMSDGVVRVAPRENLMFLDLLKLSSTTCSVCSDDPPKDPVVTLCGHVFCYECVS
VNINGDNNTCPALNCHSOLKHDVVFTESAVRSCINDYDDPEDKNALVASRRVYFIENPSC
DRDSSVACRARQSRHSTNKDNSISGLNLIFTFLKDKCNDYETGAMLSLKAGNLGLNMVA
ASHVILLDLWNPTTEDQADRAHRIGQTRAVTVTRIAIKNTVEERILTLHERKRNIVAS
ALGEKNWQKFCDSNTNRRSRISVFWCWEYPRVFDKRNKTFSYLISHKCECNE*

>G1586 (1..807)

ATGAATCAAGAAGGTGCTTCACATAGCCCATCCTCCACTTCCACCGAACAGTCCGGGCA
CGTTGGTCACCTAAACCGGAGCAAATCTTGATACTCGAATCCATCTTCAACAGTGGTACT
GTTAACCCACCAAAGATGAAACGGTGAGGATAAGAAAGATGCTTGAGAAATTCGGTGCT
GTGGGAGACGCAAACGCTCTTCTACTGGTTTCAAAACCGACGGTCAAGATCTCGCCGGAGA
CACCGGCAGCTTTTAGCAGCCACCACCGCAGCCGCCACCTCCATAGGAGCTGAAGACCAC
CAGCATGACGGCCATGAGCATGCATCAATATCCTTGACGCAACAACGAGATTGATTG
GGGTTTGAAGTTGTAGCAACTTATCAGCTAATTACTTCTTAAATGGATCGTCGTCTCT
CAATCCCTTCTTTTTCTCGGCCTCTCTTCTTCAAGTGGTGGGTGTGAGAACAACAAT
GGTATGGAGAATCTCTTCAAAATGTATGGCCATGAATCTGATCATAATCATCAGCAGCAG
CATCATAGCTCAAATGCTGCATCAGTTTTAAACCCATCTGATCAAACTCCAACCTCCCAA
TACGAACAAGAAGGGTTTATGACGGTGTTTATAAACGGAGTTCCTATGGAAGTAACAAAA
GGAGCAATAGACATGAACAAATGTTCCGGTGATGATTCCGGTGTACTTCATTCTCTGGT
CTTCTCTTCCACTGATGAGTTTGGTTTCTTGTATGCATTCTTTACAACATGGACAACT
TATTTCTGGTACCGAGACAGACATGA

>G1586 Amino Acid Sequence (domain in AA coordinates: 21-81)
MNQEGASHSPSSTSTEPVRARWSPKPEQILILESIFNSGTVNPPKDETVRIRKMLEKFGA
VGDANVIFYWFQNRSSRRRRHQLLAATTAATSIGAEDHQHMTAMSMHQYPCSNNEIDL
GFGSCSNLSANYFLNGSSSQIPSFLLGLSSSSGGCENNNNGMENLFKMYGHESDHNHQQQ
HHSSNAASVLNPSDQNSNSQYEQEGFMTVFINGVPMVETKGAIDMKTMFGDDSVLLHSSG
LPLPTDEFGFLMHSLOHGQTYFLVPRQT*

>G1786 (1..1170)

ATGATCGTGTACGGTGGGGGAGCATCCGAGGACGGTGAAGGTGGAGGGGTGGTTCCTGAAG
AAAGGGCCATGGACGGTGGCCGAGGACGAGACACTGGCGGCTTACGTACGGGAATACGGT
GAAGGGAACTGGAATTCTGTTCAGAAGAAGACATGGCTGGCTAGGTGTGGCAAGAGCTGC
CGCCTCCGCTGGGCTAACCACTTACGACCTAATCTCAGGAAAGGCTCCTTACCCCCGAC
GAAGAACGTCTCATCATACAACCTCCACTCTCAGCTAGGCAACAATGGGCTCGCATGGCT
GCTCAGTTACCAGGCAGAACAGATAACGAGATCAAGAACTACTGGAACACGAGGTTGAAA
CGCTTCCAACGCCAAGGCCTCCCTCTCTACCCTCCAGAATATTTCCAAAAACAATCATCAA
CAACAAATGTATCCTCAACAGCCCTCCTCACCTCTCCCGTCCCAACACCTGCTTCTTCC
TTTACCTTTCTCTCTCTCAACCGCCTTCTCTGTGTGCCAAACGTTGTTATAACACTGGC
TTCTCTCCCAAGGCCTCATATATTTCTCTCCAACCAATTTCTTTGTCTCGTCTCCGAC
TTCTTTCACACCAATTCCTCTCTTTCTCTCTATCAGTCTACCAATCCGGTTTACTCCATG
AAACATGAGCTCTCTTCAAACCAATTCATACTCTGCTCTTTAGGAGTCTATCAAGTA
AGCAAGTTCTCAGACAATGGGGATTGTAACCAAAACCTGAACACCGGTTTGCATACAAAT
ACCTGTCTCAGCTGTTAGAGGATCTTATGGAGGAGGCCGAGGCTCTAGCTGATAGCTTTTCGT
GCTCCTTAAGCGGAGACAAATCATGGCTGCGCTTGAGGACAACAACAACAACAACATTT
TTCTCGGGAGGTTTCGGACATCGTGTCTTCTTCCAACAGTCTATGTTCTTCAAGGTTTA
ACACCAAAGGAAGATGAGTCTCTCCAGATGAACACAAGTACAGGACATAACAAG
CTTCTTGACTGGGGAGTGAAGTGAAGAAATCTCAAACGGGCAATCTCTGTGATAACA
ACAGAGAACAACCTTGTCTTTCAGTACACCAAGTTCGCTTTTCTGTTTCCAGTTGATGAT
GACACCAACAACCTTGCCAGGGATCTGCTAG

>G1786 Amino Acid Sequence (domain in AA coordinates: TBD)

>G1786 Amino Acid Sequence (domain 1) in *Escherichia coli*
 MIVYGGGASEDEGGGVVLKGPWTVAEDETLLAAYVREYGEQWNSVQKKTWLRARCGKSC
 RLRWANHLRPNLRKGSFTPEERLIIQLHSQLGNKWARMQAQLPGRTDNEIKNYWNTRLK
 RFQRQGLPLYPPEYSQNNHQQQMYPQQPSSPLPSQTPASSFTFPLLPSPSLCPKRCYNTA
 FSPKASYISSPTNPLVSSPTFLHTHSSLSYQSTNPVYSMKHELSSNQIPYASASLGYYQV
 SKFSDNGDCNQNLNTGLHTNTCQLLEDLMEEAEALADSFRAPKRRQIMAALEDNNNNNNF
 FSGGGFGRHVSSNSLCSLQGLTPKEDESLQMNTMQDEDTIKLLDWGSEESIENGQSSVIT
 TENNLVLDDHQAFLFPVDDDTNNLPGIC*

>G1792 (77..496)

AATCATGATGCTCTTATTAAATAACAGTGTCTGACCAAGCTCTTACAAAGCAAACCAATC
 TAGAACACCAAAGTTAATGGAGAGCTCAAACAGGAGCAGCAACAACCAATCACAAGATGA
 CAAGCAAGCTCGTTTCCGGGGAGTTCCAAGAAGGCCTTGGGGAAAGTTTGCAGCAGAGAT
 TCGAGACCCGTCGAGAAAACGGTGCCCGTCTTTGGCTCGGGACATTTGAGACCCGTCGAGGA
 GGCAGCAAGGGCTTATGACCCGAGCAGCTTTAACTTTAGGGGTATCTCGTATACTCAA
 CTTCCCTTAATGAGTATATCACGTCATGACGACTACTCGCTTCGCCCTCCTTATGCTTC
 TTCTTCTTCGTCGTCGTCATCGGGTTCAACTTCTACTAATGTGAGTCGACAAAACCAAAG
 AGAAGTTTTCGAGTTTGAGTATTTGGACGATAAGGTTCTTGAAGAACCTTCTTGATTGAGA
 AGAAAGGAAGAGATAATCACGATTAGTTTGTGTTTGTATTTTATGTGGCACTGTTGTGTG
 CTACCTACGTGCATTATGTGTCATGTATAGGTCGCTTGATTAGTACTTTATAACATGCATG
 CCACGACCATAAAATTGTAAAGAGAAGAGCTACTTTGCGTTTTCATGAAATATGAATGTTAG
 ATGGTTTGTAGTACAAAAA

>G1792 Amino Acid Sequence (domain in aa coordinates: 17-85)
MESSNRSSNNQSQDDKQARFRGVRRRPWGKFAAEIRDPSRNGARLWLGTFTAEAAARAY
DRAAFNLRGHLAILNFPNEYYPMDYSLRPPYASSSSSSSSSGSTSTNVSQRNQREVFEE
EYLDKVLVEELLDSSEERK*

>G1865 (48..899)

AAGAAGAGGACATGAAGCACAGAGATTCTGCAGACTGCAGGTGACCAATGGACACTTTAT
CAATAAAAAACATACTACTACTCTCTTACACTTTCAATTTTCCAATACAAATCCCAATCT
TTAATCTCTCTTTCTTCTTCATCTCTCTTTCTCTTTCTCTTTCATGGCTACAAGGATTC
CATTACAGAATCACAATGGGAAGAAGTTGAAAACCAAGCTCTTGTGTTCAAGTACTTAG
CTGCAAAATATGCCTGTTCCACCTCATCTTCTCTTCTCTCAAAAAGACCCCTTTCTCTTCT
CTTCTTCTTCTTCTTTCATCTTCTTCTTCAAGCTTCTTCTCTCCCACTCTTTCTCCACACT
TTGGGTGGAATGTGTATGAGATGGGAATGGGAAGAAAGATAGATGCAGAGCCAGGAAGAT
GTAGAAGAACTGATGGCAAGAAATGGAGATGCTCTAAAGAAAGCTTACCCTGACTCTAAGT
ACTGTGAGAGACATCATGATGAGGCCAAGAACCGTTCTTCTCAAGAAGGCCCTCCTCCTA
CTCAATTACTCTCAAATCTCTTTCTCGACTCTTCTTCCAGAAGAAGAAGTGATACATA
TGGATGATTCTTCTTCCATAGAACCTTCCGGGTCAATCAAAGCTGCTCTGGCTCAGCAA

TGGAAGATAATGATGATGGCTCATGTAGAGGCATCAACAACGAGGAGAAGCAGCCGGATC
GACATTGCTTCATCCTTGGTACTGACTTGAGGACACGTGAGAGGCCATTGATGTTAGAGG
AGAAGCTGAAACAAAGAGATCATGATAATGAAGAAGAGCAAGGAAGCAAGAGGTTTATA
GGTTTCTTGATGAATGGCCTTCTTCTAAATCTTCTGTTTCTACTTCACTCTTCATTTGAT
CATCTTTTGTCTTTATAACCTTGTATTTCTTGTTAAGATGGTAATGCAAATT
>G1865 Amino Acid Sequence (domain in AA coordinates: 124-149)
MDTLSIKTYLLLSYTFNFIPIPIFNLSFFFISLSLSLFMATRIPFTESQWEELENQALV
FKYLAANMPVPPHLLFLIKRPFLFSSSSSSSSSSSFSPTLSPHFGWNVYEMGMGRKIDA
EPGRCRRTDGKKWRCSKEAYPDSKYCERHMRGKNRSSSRKPPPTQFTPNLFLDSSSRRR
RSGYMDDFFSIEPSGSIKSCSGSAMEDNDDGSCRGINNEEKQPD RHCFILGTDLRTRERP
LMLEEKLRQDRHDNEEEQSKRFYRFLDEWPSSKSSVSTSLFI*
>G1886 (43..909)
AGGAAACATAAGTAATCGTTGCTTCGATCCTTTGTTACATGGATGGATCCTGAACAGGAA
ATCTCAAACGAGACTTTGGAACTATATTGGTAAGTTCAACAAAAGGAAGCAATAATAAC
AATAAGAAAATGGAAGAAGAAATGAAGAAGAAAGTATCAAGAGGAGAATTAGGAGGTGAA
GCTCAAAATTGTCCAAGATGTGAATCTCCAAACACAAAGTTTTGTTACTACAACACTAT
AGTCTCTCACAACTCGTTACTTCTGCAAACTTGTGCGGAGATATTGGACTAAAGGCGGT
ACTCTTCGTAACGTTCCCGTCGGTGGTGGTTGCCGTCGAAACAAACGATCCTCTTCCTCA
GCTTTCTCCAAGAACAACAATAAGTCTATTAATTTCCATACTGATCCACTTCAGAAC
CCTTTAATTACGGGAATGCCACCATCATCTTTTGGTTATGATCACTCCATTGATCTCAAC
CTCGCTTTTCGCTACTCTCCAAAGCATCATTTATCCTCTCAAGCTACTACGCCTTCTTTT
GGGTTTGGAGGTGATCTTTCTATTTATGGAACTCAACGAATGATGTAGGGATCTTCGGA
GGGCAAAACGGTACTTATAACAATAGTTTGTGTTATGGGTTTATGTCCGGAATGGTAAT
AATAATCAAAATGAAATCAAGATGGCTTCTACATTGGGGATGTCTTTGGAAGGAAACGAG
AGAAAGCAAGAGAATGTGAACAATAACAATAAATACTCAGAGAATCCTAGCAAGGTGTTC
TGGGGGTTTCCATGGCAGATGACCGGAGATTCCGCCGGAGTTGTACCGGAGATTGATCCC
GGAAGGGAAAGCTGGAATGGGATGGTTTCATCTTGAATAATGGTTTACTCAACACTCCT
TTGGTCTAGCAGATCATTA
>G1886 Amino Acid Sequence (domain in aa coordinates: 17-59)
MDPEQEISNETLETILVSSTKGSNNNNKMEEMKKKVS R GELGGEAQNCP RCESPNTKF
CYNNYSLSQPRYFCKSCRRYWT KGGTLRNV PVGGGCRNRKRSSSAF SKNNNNKS INFH
TDPLQNPLITGMPPSSFYDHSIDLNLAFATLQKHHLSSQATTPSF GFGDLSIYGNSTN
DVGI FGGQNGTYNNSLCYGFMSGNGNNQNEIKMASTLGMSLEGNERKQENVNNNNNNSE
NPSKVFWGFPWQMTGDSAGVVPEIDPGRESWNGMVSSWNNGLLNTPLV*
>G1933 (33..1418)
AATTGAGATTAAAGTAATTTATCTTTTCAGAAAATGGCGGTTGAAGACGATGTATCTTTGA
TAAGAACGACGAGCTTAGTGCCACCAACAAGACCCACGATTACAGTTCTCATAGACCTC
CGGCGATCGAAACGGCGGCGTATTTCTTTGGCGGTGGAGATGGGCTTAGTCTAAGCCAG
GGCCACTTTCTTTTGTCTCTTCTTTGTTTGTGATAACTTCCCTGACGTCTTGACGCCG
ATAACCAACGGACGACGTCTTTACTCAGCTTCTTAACGGAACATGTGCGGTGTCTCCTG
GTGGCGGAGGACGTTCAACGGCGGGGATGTTCCGCCGGAGGAGTCCGATGTTTACAATCC
CTTCTGGTTTCAGCCCTTCTAGTCTTCTCACCTCGCCCATGTTCTTTCCCCCGAGTCGT
CAGCTCATACCGGCTTTATTC AACACGGCAGCAGTCACAACCGCAACCACAACGACCAG
ACACGTTTCTCTACCATATGCCACCATCGACATCCGTGCGCGTCCATGGTGTCAATCTT
TAGACGTTTTCACAGTAGATCAAAGAGCTCGAAACCATTATAATAATCCGGGAATAACA
ATAATAACCGGTCGTATAACGTTGTGAACGTTGATAAACGGCGGATGACGGTTATAACT
GGAGGAAGTACGGACAAAAGCCTATCAAAGGGTGTGAATATCCAAGGAGTTATTACAAAT
GTACACATGTTAACTGTCCGGTGAAGAAGAAAGTCAACGGTCATCGGATGGACAGATCA
CTCAGATCATTTACAAAGTCAACATGATCACGAGAGGCCTCAGAATCGCCGTGGCGGTG
GAGGCAGAGATTCCAGTGAGGTTGGTGGTGCAGGGCAAATGATGGAATCTAGTGATGATA
GTGGTTATCGTAAGGATCATGATGATGATGATGATGATGAAGATGATGAAGATCTTC
CGGCTTCAAAGATAAGAAGAAATAGACGGTGTGTGACGACTACCGGACGGTGACCGAGC
CTAAGATTATCGTTTCAGACAAAAGTGAAGTCGATCTTCTCGACGATGGCTATAGGTGGC
GTAAGTACGGACAAAAGTTGTCAAAGGAAATCCCCATCCAAGGAGCTATTATAAATGTA
CAACGCCAAATTGTACGGTCCGTAAACATGTAGAGAGAGCTTCCACGGATGCTAAGGCTG
TGATTACAACCTACGAAGGTAACACAATCACGATGTCCCTGCCGCTAGAAACGGTACCG
CGGCAGCAACCGCAGCTGCGGTGGGGCCGTCTGACCACCATCGTATGAGATCAATGTCCG

GGAACAATATGCAACAACATATGAGTTTCGGTAACAATAATAACACAGGCCAATCTCCGG
TTCTTTTGAGGTTGAAAGAAGAGAAAATCACAATTTGACTTTTAAGAACCAAAGATTTCG
AGATTGATATT

>G1933 Amino Acid Sequence (conserved domain in AA coordinates:205-263, 344-404)

MAVEDDVSLIRITTLVAPTRPTITVPHRPPAIETAAYFFGGGDGLSLSPGPLSFVSSSLFV
DNFPDVLTPDNQRTTSFTQLLNGTMSVSPGGGRSTAGMFAGGGPMFTIPSGFSPSSLLT
SPMFFPPQSSAHTGFIQPRQSQPQRPDTFPHMPPSTSVAVHGRQSLDVSQVDQRRAR
NHYNPNPNNNNNRSYNVVNVDPADGYNWRKYGQKPIKGCEYPRSYKCTHVNCVPVKKK
VERSSDGQITQIIYKQHDHERPQNRRCGGGRDSTEVGAGQMMESSDDSGYRKDHDDDD
DDDEDDDELPAKIRRIDGVSTTHRTVTEPKIIVQTKSEVDLLDDGYRWRKYGQKVVKGN
PHPRSYYKCTTPNCTVRKHVERASTDAKAVITTYEGKHNHDVPAARNGTAAATAAAGVPS
DHHRMRSMSGNNMQQHMSFGNNNNTGQSPVLLRLKEEKITII*

>G2059 (58..1089)

TTAAGAACAGGCTTCATTCTCTGGACAAACACTCAAAAAACAAACAAAAAAGGAACATG
GAAGATCAGTTTCTTAAATAGAACTAGCTTCATGCACGACAAGCTCTTGTCTTCTGGA
ATCTACGGGTTCTTGAGTTCTTCGACGCCGCCACAACCTCTCGGTGTTCCAATATTTTG
GAAGGTATGAAATCTCTCTTCTTCTCTGCTTCTTCGACTCCGAGCTACTTTGTGTGCGCT
CATGATCATGAGCTCACATCTTCTATTCATCCATCTCCGGTAGCTTCTGTTCTTGGAAC
TTTCTAGAATCTTTCTCAGTCTCAACATCCTGATCATCATCTTCTAAACCTCCAAAC
CTTACTTTGTCTCTTAAAGAACCAAGCTACTAGAACTTTCTCAATCCGAAAGCAACATG
AGCCCTTACCATAAATACATCCCAACTCTTTTATCAATCAGACCAAAACAGAAACGAA
TGGGTAGAGATCAATAAACTCTAACCAACTATCCCTCGAAAGGTTTGGAACTATTGG
CTAAGTACCACCAAGACTCAACCCATGAAGTCAAAAACAAGAAAGGTTGTTAGACGACG
ACCCCAACAAAACGTGTATAGAGGAGTGAGACAAAGACACTGGGGCAAATGGGTGCGAGAG
ATTAGGCTTCCAAGGAACAGAACCCGTGTTTGGCTCGGCACCTTTGAAACCGCTGAGCAA
GCAGCAATGGCTTACGATACAGCAGCTTATATCCTTCGTGGCGAATTGCGACACCTCAAC
TTTCTGATCTTAAACACCAGCTCAAGTCCGGTCTTTGCGATGCATGATCGCCTCACTT
TTGGAGTCCAAGATTCAACAGATCTCATCTTCCCAAGTAAGTAACCTCTCCTTCTCCTCCT
CCTCCAAAAGTGGGAACACCGGAGCAAAGAATCATCATGAAGATGGAGTCAGGAGAA
GACGTGATGATGAAGAAACAGAAAAGCCATAAGGAAGTGATGGAAGGAGATGGTGTACAA
TTGAGTAGGATGCCTTCTTTGGATATGGATCTCATTTGGGATGCTCTCTCATTTCTCAT
TCTTCTTGACTTCAAATTAATATTTGTCAAACCTATTTTACTTACTTCTACCCCTTTTATA
TATCAAAAGTTTCCACCAAGAAAGAAATTATATTATGATGCCAAGATTGGTTTGCATT
TGGGGTTGAACACATTGTAATCTTCTTACGACCACATAATCAAGTGGTCTCTCTTTT
TGTCTGCTAA

>G2059 Amino Acid Sequence (conserved domain in AA coordinates:184-254)

MEDQFPKISTFSMHDKLLSSGIYGLSSSTPPQLLGVPFLEGMKSPLLPASSTPSYFVS
PHDHELTSSIHPSVPASVPWNFLESFPQSQHPDHHPKPPNLTFLKEPKLLELSQSES
MSPYHKYIPNSFYQSDQNRNEWVEINKTLTNYPKSGFGNYWLSTTKTQPMKSKTRKVQT
TPTTKLYRGVRQRHWGKWVAEIRLPRNRTRVWLGTFTAEQAAMAYDTAAYILRGEFAHL
NFPDLKHQLKSGSLRCMIASLLESKIQQIISSSQVSNSSPSPPPKVGTPQKNHMKMESG
EDVMMKKQKSHKEVMEGDGVQLSRMPSLDMDLIWDALSFPHSS*

>G2105 (42..1487)

CTCTCTGACTTGAACCTTCTCTTCTACCGAATCAAACCAAATGGAGGATCATCAAAC
ATCCACAGTACGGTATAGAACAACCATCTTCTCAATTCTCCTCTGATCTCTTCGGCTTCA
ACCTCGTTTCAGCGCCGGACCAGCACCATCGTCTTCATTTACCGACCATGAGATAAGTT
TATTGCCACGTGGAATACAAGGGCTTACGGTGGCTGGAAACAACAGTAACACTATTACAA
CGATCCAGAGTGGTGGCTGTGTTGGTGGGTTTAGTGGCTTTACGGACGGCGAGGAACAG
GGAGGTGGCCGAGGCAAGAGACGTTGATGTTGTTGGAGGTCAGATCTCGTCTTGATCACA
AGTTCAAAGAAGCTAATCAAAGGGTCCTCTCTGGGATGAAGTTTCTAGGATTATGTCGG
AGGAACATGGATACACTAGGAGTGGCAAGAAGTGATAGAGAGAAGTTCGAGAATCTCTACA
AGTACTATAAAAAACAAAAGAAGGCAAATCCGGTCGGCGACAAGATGGTAAAAACTATA
GATTTTCCCGCAGCTTGAAGCGATATACGGCGAATCCAAAGACTCGGTTTCTTGCTATA
ACAACACGCAGTTTCAATGACCAATGCTCTTCATAGTAATTTCCGCGCTTCTAACATT
ATAACATCGTCCCTCATCATCAGAAATCCCTTGATGACCAATACCAATACTCAAAGTCAA
GCCTTAGCATTTCTAACAATTTCAACTCCTCCTCCGATTGGATCTAATCTTCTCTCTG
AAGGAAACGAACTACTAAAAGAGAGGGGATGCATTGGAAGGAAAAGATCAAGGAATTCA

TTGGTGTTCATATGGAGAGGTTGATAGAGAAGCAAGATTTTTGGCTTGAGAAGTTGATGA
AGATTGTGGAAGACAAAGAACATCAAAGGATGCTGAGAGAAGAGGAATGGAGAAGGATTG
AAGCGGAAAGGATCGATAAGGAACGTTCTGTTTTGGACAAAAGAGAGGGAGAGGATTGAAG
CTCGGGATGTTGCGGTGATTAATGCCTTGACAGTACTTGACGGGAAGGGCATTGATAAGGC
CGGATTCTTCGCTCTCTACAGAGAGGATTAATGGGAATGGAAGCGATAAAATGATGGCTG
ATAATGAATTTGCTGATGAAGGAAATAAGGGCAAGATGGATAAAAAACAAATGAATAAGA
AAAGGAAGGAGAAATGGTCAAGCCACGAGGGAATCATCCAAGAACCAAAGAGAATATGA
TGATATACAACATCAAGAACTAAGATTAATGATTTTTGTGCGAGATGATGACCAATGCC
ATCATGAAGGTTACTCACCTTCAAACCCAAGAACGCAGGAAGTCCGAGCTGCAGCAATG
CCATGGCAGCTAGTACAAAGTGCTTTCCATTGCTTGAAGGAGAAGGAGATCAGAAGTTGT
GGGAGGGTTATGGTTGAAGCAAAGGAAAGAAAATAATCATCAGTAAGCTACATTTTTCA
TTCTCAAATGAAGAATAAGAGAAGTTAGAAAACGAT

>G2105 Amino Acid Sequence (domain in AA coordinates: 100-153)

MEDHQNHPPQYGIEQPSSQFSSDLFGFNLVSAPDQHRLHFTDHEISLLPRGIQGLTVAGN
NSNTITTIQSGGCVGGFSGFTDGGGTGRWPRQETLMLLEVRSLDHKFKEANQKGPLWDE
VSRIMSEEHGYTRSGKKCREKFENLYKYYKKTKEGKSGRRQDGKNYRFFRQLEAIYGESK
DSVSCYNNQTQFIMTNALHSNFRASNIHNI VPHQNPLMTNTNTQSQSLISNNFNSSDL
DLTSSSEGNETTREGMHWKEKIKEFIGVHMERLIEKQDFWLEKLMKIVEDKEHQRLRE
EEWRRIEAEIRIDKERSFWTKERERIEARDVAVINALQYL TGRALIRPDSSSPTERINNG
SDKMMADNEFADEGNKGMKMDKKQMNKKRKEKWSSHGGNHPRTKENMMIYNNQETKINDFC
RDDDQCHHEGYSPSNSKNAGTPSCSNAMAASTKCFPLLEGEQDNLWEGYGLKQRKENNH
Q*

>G2117 (49..465)

ATACTTGTCACAAAAAATTTCTTAAAGAACGCATAACTGTTTTTTTCATGGCTGGTTCT
GTCTATAACCTTCCAAGTCAAAACCTAATCCACAGTCTTTATCCAAATCTTTGTTGAT
CGAGTACCACTTTCAAACCTTGCCCTGCCACGTGAGAGGACTCTAGCCGAGTGCAGAAGAT
AATGAGAGGAAGCGGAGAAGGAAGGTATCGAACCGGAGTCAGCTCGGAGATCGCGTATG
CGGAAACAGCGTCACATGGAAGAACTGTGGTCCATGCTTGTTCAACTCATCAATAAGAAC
AAATCTCTAGTCGATGAGCTAAGCCAAGCCAGGGAATGTTACGAGAAGGTTATAGAAGAG
AACATGAAACTTCGAGAGGAAAACCTCAAGTCGAGGAAGATGATTGGTGAGATCGGGCTT
AATAGGTTTCTTAGCGTAGAGCCGATCAGATCTGGACCTTCTAATCGTCTCGTAAGCTT
GTTGGTTTTTTGTTGTTTATTAAAG

>G2117 Amino Acid Sequence (conserved domain in AA coordinates: 46-106)

MAGSVYNLPSQNPNPQSLFQIFVDRVPLSNLPATSDSSRTAEDNERKRRRKVSNRESAR
RSRMQRQHMEELWSMLVQLINKNKS LVDLSQARECYEKVIEENMKLRENSKSRKMIG
EIGLNRFLSVADQIWF*

>G2124 (87..923)

GAACAGCAAAACCTAGATTTCTGTTCAGCTCAAGACCGTACAAAACCTTGGAAGTCA
TATATAAGATCTCGAGAATAGCATTATGAATATCGTCTCTTGGAAGATGCAAACGACG
AAGTTGCAGGCGGCGCTACGACAAGACGTGAAAGAGAAGTAAAGAGGATCAAGAAGAAA
CCGAAGTCAGAGCCACCAAGTGGCAAAACCGTAATTAAGAGCAGCCTACATCGATCTCTT
CTTCTTCTTCTTCTGTTGATGAAATCCAAGGATCCGAGGATTGTTAGGGTTTCACGCGCCT
TTGGAGGCAAGACCGTACAGCAAAGTGTGTACGTTACGTGGACTACGTGACAGACGCG
TGAGATTATCAGTCCCAACGGCTATTACAGCTCTACGATCTTCAAGAACGGCTCGGTGTTG
ACCAGCCTAGCAAAGCCGTTGACTGGTTGCTTGATGCAGCTAAAGAGGAGATCGACGAGC
TACCTCCGTTACCTATCTCGCCGAAAATTTACGATCTTCAACCATCATCAGTCCTTCT
TGAATCTTGGTCAACGGCCCGGTCAAGATCCGACCCAACCTCGGGTTTAAATCAATGGAT
GTGTACAAAAGTCTACTACTACTAGCCGGAAGAAAACGATAGAGAGAAAGGAGAAAACG
ATGTCGTTTACACAAACAATCATCATGTTGGGTCTTATGGAAGTTATCACAACTGGAAC
ATCATCATCATCATCAACAATTTGAGTTTACAGGCAGATTATCATAGTCATCAACTAC
CTACAACATATACAATCTTTGTTTCCATCATCATCGTCAGCTGGTTTCAGGGACTATGGAGA
CATTAGATCCGAGGCAAAATGTAGCAACAATGGTGGTAGAGACATTGATAATCGGATGTG
TCGGTCCAATTCACCGAACTAATAGCACTACAACGGCTAACATGTGAGGCATCTAGGC
TCGGAGCGTTGTACAAGTAGAGGAAGTGATCACCATATGTGAAGTTAGATTATTGAAACG
ATATAATTGTTGTTGATGTGTTTCAGAAATAAGGGGACAC

>G2124 Amino Acid Sequence (domain in AA coordinates: 75-132)

MNIVSWKDANDEVAGGATRRREREVKEDQEETEVRATSGKTVIKKQPTSISSSSSSWMKS
KDPRIVRVSRAFGGKDRHSKVCTLRGLRDRRVRLSVPTAIQLYDLQERLGVDQPSKAVDW
LLDAAKEEIDELPPLPISPENFSIFNHHQSFLNLGQRPQDPTQLGFKINGCVQKSTTTS
REENDREKGENDVVYTNHHVGSYGTYNLEHHHHHHQHLSLQADYHSHQLHSLVPFPSQ
ILVCPMTTSPTTTTIQSLFPSSSSAGSGTMETLDPQOM*

>G2140 (148..1254)

ACTCTCTTAACCTTTTCGTTTCTTCTCCTACCTTCTTTTACCAACCTTTCTCTTCTTACA
CACATATATATATACATATATAGAGAGAGAGAAGAGGACAAAGAGTTGAAAGATGAAGAC
TCTCATGTCTTCATAGAAACAAGTGATATGTGCGCTAAGAAAGAAGAAGAAGAAGAA
GAAGAAGACAGTTCTGAAGCCATGAACAACATACAAAATTACCAAATGACCTCTTCTTT
CACCAACTCATCTCTCATCATCACCATCATCATCATGATCCTTCTCAATCTGAAACTTTG
GGAGCATCCGGTAACGTTGGATCTGGTTTCACTATCTTCTCTCAAGATTCCGTCTCTCCA
ATATGGTCTCTACCTCCACCTACCTCGATCCAACCACCATTGATCAGTTTCTCTCTCTCT
TCTTCTTCTCCAGCATCTTCTACGGAAGTTTCTTCAACAGAAGTCGAGCTCATCATCAG
GGATTACAGTTTGGGTACGAGGGTTTTGGTGGAGCCACGTGAGCAGCACATCATCATCAT
GAACAACCTTCGGATCTTGTGCGGAAGCTTTAGGTCCGGTAGTACAAGCCGGGTCCGGTCT
TTTGGGTTACAAGCTGAGTTAGGGAAGATGACAGCACAAGAGATCATGGACGCTAAAGCT
TTGGCTGCTTCAAAGATCATAGTGAAGCTGAGAGAAGAAGAAGAGAGAGAATCAATAAT
CATCTCGCTAAGCTCCGTACGATATTACCCAACACCACCAAAACGGATAAAGCGTCGTTA
CTAGCTGAAGTGATCCAACATGTGAAAGAGTTGAAGAGAGAGACTTCAGTGATCTCAGAG
ACAAATCTTGTCCCAACGGAAAGCGATGAGTTAACGGTAGCTTTCACGGAGGAGGAAGAA
ACCGGAGATGGCAGATTTGTAATTAAGCGTCGCTTTGCTGTGAAGACAGGTCGGATCTC
TTGCCGTGACATGATTAAAAACATTGAAAGCTATGCGTCTCAAACGCTCAAGGCGGAGATA
ACCACCGTTGGGGGACGAGTCAAGAACGTTTTGTTTGTACCGGAGAAGAGAGCTCCGGT
GAGGAAGTGAGGAAGAGTACTGTATAGGGACGATTGAGGAAGCTTTGAAAGCGGTGATG
GAGAAGAGCAATGTAGAGGAATCATCTTCTTGGAAATGCTAAGAGACAGAGAATGAGT
AGTCACAACACTATCACTATCGTCAACAACAACAACAATATAATCAGAGGTAATCAATT
TTTTACTTAAATCGCTTTTTTTTCTTACTTTCGGTGTATCTACTACGTGTGTTGTTTGTCT
GGTTATGGAAATGAATGTTGTACGTACGTTATACTATAGATATATGTGTGTTTGTGTGT
ATGTATAACGGAAGTATTTGTATCCGTTGTGGTCTTGGACTTTTGGTTTGGTTCTAAGAT
ACTTATTTTAAAAACCTGTATCGTTGAGTTGGTTTTCTAGATATGCTTAATGGGAGTAT
GTGACGAAAAAAA

>G2140 Amino Acid Sequence (domain in AA coordinates:167-242)
MCAKKEEEEEEDSSSEAMNINQNYQNDLFPHQLISHHHHHHHHDPSQSETLGASGNVSG
FTIIFSQDSVSPISLPPPTSIQPPFDQFPPSSSPASFYGSFFNRSRAHHQGLQFGYEGF
GGATSAHHHHEQLRLILSEALGPVVQAGSGPFGLOAELGKMTAQEIMDAKALAASKSHSE
AERRRRERINHLAKLRSILPNTTKTDKASLLAEVIQHVKEKRETSVISETNLVPTESD
ELTVAFTEEEETGDGRFVIKASLCCEDRSDLLPDMIKTLKAMRLKTLKAEITTVGGRVKN
VLFVTGESSSGEEVEEYICGTIEBALKAVMEKSNVEESSSSSGNAKRQRMSSHNTITIVE
QQQQYNQR*

>G2144 (102..1241)

ATTAGGTTTGTGTCGTGAGATTTGATTACACAAATTGCTGAATTTGGTTTCGATTAT
TGGTGTATTTGTTTTCGAAGATTTCCAGTGAGTTTCCGTTTATGGATCTGACTGGAGGAT
TTGGAGCTAGATCCGGCGGTGTTGGACCGTGCCGGGAACCAATAGGCCTTGAATCGCTAC
ATCTCGGTGACGAATTTCCGCAACTAGTGACGACTTTACCTCCCGAGAACCCCGGCGGTT
CGTTCACGGCTTTGCTTGAGCTTCCACCTACACAAGCAGTGGAGCTTCTCCATTTCACTG
ATTCTTCGTCTTCTCAACAAGCGGCAGTGACAGGGATCGGTGGAGAGATTCCTCCGCCGC
TTCACTCTTTCGGTGGGACATTGGCTTTTCTTCTAACTCAGTTCTCATGGAGCGAGCAG
CTCGTTTCTCGGTGATTGCCACTGAGCAACAAAACGGAAATATCTCCGGGGAGACTCCGA
CGAGCTCTGTACCTTCAATTCAAGTGCTAATCTCGACAGAGTCAAGACGGAGCCTGCTG
AGACCGATTATCTCAGCGGTTGATTTCTGATTACGCGATTGAGAATCAAATCCCTTGCC
CTAACCAGAACAATCGAAATGGGAAGAGGAAAGATTTCGAAAAGAAGGTTAAAAGCTCGA
CGAAGAAGAACAAGCTCTGAAGAGAACGAGAAGCTGCCATATGTTACGTTAGAGCTC
GTCGTGGTCAAGCAACCGATAGCCATAGCTTAGCAGAACGAGCAAGAAGAGAGAAGATAA
ATGCACGAATGAAGCTGTTACAGGAACGTGCCAGCTGTGATAAGATTCAAGGTACCG
CGCTGGTGTGATGAAATCATTAACCATGTCCAGTCATTACAACGCTCAAGTGGAGATGC
TATCAATGAGACTTGCTGCGGTAAACCCAGAATCGACTTCAATCTCGACACCATATTGG

CTTCAGAAAACGGTTCTTTAATGGATGGGAGCTTCAATGCCGCACCAATGCAGCTTGCTT
GGCCTCAGCAAGCCATTGAGACCGAACAGTCCTTTTCATCACCGGCAACTGCAACAACCAC
CAACACAACAATGGCCTTTTGACGGCTTGAACCAGCCGGTATGGGGAAGAGAAGAGGATC
AAGCTCATGGCAATGATAACAGCAATTTGATGGCAGTTTCTGAAAATGTAATGGTGGCTT
CTGCTAATTTGCACCCAAATCAGGTCAAAATGGAGCTGTAAGTTGGGAAAACGGTAGAGA
TCATGAATGTGTATATACATCGTATAAGCTCGTTTCTCTCTATATAAATATAATCATAAA
TATAGATATCTGTTAAGAAGGTATCAGTCATTTGATTAGAGAGACAACACTGGTATGAT
TGTTTCTTATTCTTGTACCAGATTCGACAATGTAGAATTTAGTAGGATATGATCATT
GATCTCGTTATATATA

>G2144 Amino Acid Sequence (domain in AA coordinates:203-283)
MDLTGGFGARSGGVGPCREPIGLES LHLGDEFRLVTTLPENPGGSFTALLELPPTQAV
ELLHFTDSSSSQQA AVTGIGGEIPPLHSFGGTLAFPSNSVLMERAARFSVIATEQQNGN
ISGETPTSSVPSNSSANLDRVKTEPAETDSSQRLISDSAIENQIPCPNQNNRNGKRKDFE
KKGKSSTKKNKSSEENEKLPYVHVRARRGQATD SHSLAERARREKINARMKLLQELVPGC
DKIQGTALVLDEIINHVSQSLQRQVEMLSMRLAAVNPRIDFNLDTILASENGSLMDGSFNA
APMQLAWPQQAIETEQS FHRQLQPPPTQQWPF DGLNQPVWGREEDQAHGNDNSNLMAVS
ENVMVASANLHPNQVKMEL*

>G2431 (47..1057)

CCCTTTTCGTTTTTATTAAATTTCTTGGGTCGTTTCTTAAATTTGTATGTGTTTATTAAT
GGAGATCAACAATAATGCCAACAATACTACTACTATTGATAATCACAAGGCAAAGAT
GAGCCTTGTGTTGTCAACGGATGCTAAGCCAAGGTTGAAATGGACTTGTGATCTTCATCA
CAAATTCATCGAAGCCGTTAATCAACTTGGAGGACCTAACAAAGCAACACCTAAGGGTTT
GATGAAGGTTATGGAGATTCCTGGGCTTACCTTATACCATCTCAAGAGCCATTTACAGAA
ATATCGGTTAGGGAAGAGCATGAAGTTCGATGATAACAAGCTAGAAGTTTCTCTGCATC
AGAGAAATCAAGAAGTTGAGAGTAAAAACGATTCAAGAGATCTCCGAGGCTGCAGTGTAC
CGAAGAAAACAGCAATCCAGCTAAAGAAGGGCTACAAATCACAGAGGCTTTACAAATGCA
GATGGAAGTTTCAGAAGAACTTCATGAACAAATCGAAGTTCAGAGGCATTTGCAGGTGAA
GATTGAGGCACAAGGAAAGTATCTACAGTCCGTTTTAATGAAAGCTCAACAAACTCTCGC
TGGCTACTCATCTTCAAATCTCGGCATGGATTTTGGCAGGACCGAGCTCTCTAGATTAGC
TTCAATGGTGAACAGAGGCTGTCCAAGCACTTCGTTCTCAGAGCTAACGCAAGTAGAAGA
AGAAGAAGAGGTTTCTTGTGGTACAAGAAACCAGAAAACAGAGGAATTAGTCAGCTGAG
ATGTTTCAGTAGAGAGCTCGTTGACATCTTCAGAGACCTCAGAGACAAAACCTGGATACTGA
CAATAACCTTAATAAATCGATTGAACCTCCGTTGATGGAGATCAACTCGGAAGTGATGAA
GGGGAAGAAGAGAAGCATAAACGACGTCGTTTGGCTGGAGCAGCCTCTAATGAAGAGAGC
TTTTGGAGTTGATGATGATGAGCATTTGAAGTTGAGTTTGAATACTTACAAGAAAGACAT
GGAGGCGTGTACGAACATAGGACTAGGGTTTAATTAACAAAAAACAATTTTACTAAAGTT
ATATAAAATGTTTTAAAGAATCCA

>G2431 Amino Acid Sequence (conserved domain in AA coordinates:38-88)

MCLLMEINNANNTNTTIDNHKAKMSLVLSTDAKPRCLKWTCDLHHKFIEAVNQLGPNKA
TPKGLMKVMEIPGLTLYHLKSHLQKYRLGKSMKFDDNKLEVSSASENQEVESKNDSRDLR
GCSVTEENSNPAKEGLQITEALQMOMEVQKKLHEQIEVQRHLQVKIEAQGKYLQSVLMKA
QQTLAGYSSSNLGMDFARTELSRLASMVNRGCPSTSFSELTQVEEEEEGFLWYKKPENRG
ISQLRCVSESLTSSETSETKLDTDNNLNKSIELPLMEINSEVMKGKKRSINDVVCVEQP
LMKRAFGVDDDEHLKLSLNTYKDM EACTNIGLGFN*

>G2465 (86..1150)

CAATATTCTTCTCCATTGAGATTAAGCTTCTTTCTCGCTGTCGTCTCTCTATAGATCTT
GGTTCTTAGTCCCTTTTGAATAATAATGATGGTGGAGATGGATTACGCTAAGAAAATGCA
GAAATGTCATGAATACGTTGAAGCACTTGAAGAAGAACAGAAGAAAATCCAAGTCTTTCA
ACGCGAGCTTCTTTTATGTTTAGAGCTTGTCACTCAAGCGATCGAAGCTTGTGGAAGGA
GTTATCTGGTACGACGCAACTACATCAGAACAGTGTTCAGAACAGACCACAAGTGTTTG
TGGTGGTCTCGTCTTTGAAGAGTTTATTCCTATCAAGAAAATTAGTTCCTTGTGTGAAGA
AGTACAAGAAGAAGAAGAAGATGGTGAACATGAATCTTCTCCAGAACTTGTGAATAA
TAAGAAATCAGATTGGCTTAGATCTGTTTCAAGCTATGGAATCATTACCGGATCTAAATCC
AAAAGAGGAGCGTGTAGCTAAGAAAGCGAAAGTGGTGGAGGTGAAACCAAAAGCGGTGC
GTTTCAGCCGTTTCAAAAGCGCGTTTGGAGACTGATTGCAACCGGCGGTGAAAGTAGC
TAGTTCGATGCCAGCGACGACGAGTCTACGACGGAACTTGTGGTGGTAAAAGTGA
TTTGATTAAAGCTGGAGATGAGGAAAGACGGATAGAGCAGCAGCAATCGCAGTCGCATAC

GCATAGAAAACAAAGGCGGTGCTGGTCGCCGAATTACACCGTCGATTCTTAAACGCGCT
TCAGCAGCTTGGAGGATCTCATGTTGCTACACCAAGCAAATCAGGGATCACATGAAGGT
TGATGGATTAAACAAACGACGAAGTTAAAGCCATTTACAGAAATATAGACTTCACACAAG
AAGGCCAGCAGCAACATCCGTGGCGGCACAAAGTACCGGAATCAGCAACAACCACAATT
TGTGGTGGTTGGAGGCATATGGGTACCATCGTCACAAGATTTTCCACCACCGTCCGATGT
AGCCAACAAGGTGGTGTATATGCTCCGGTTGCGGTGGCGCAATCTCCAAAACGTTCCGTT
GGAGAGAAGTTGCAACTCGCCGGCGGCATCTTCTCTACAAATACAAATACTTCTACTCC
TGTGTCATAATCTGATAGTCATACTATAATCATCTCCTGATGTTGATTTTGGTGTAGGTT
TGAAAATGTTTATGTGAATGTAA

>G2465 Amino Acid Sequence (conserved domain in AA coordinates:219-269)

MMVEMDYAKMQKCHEYVEALEEEQKKIQVFQRELPLCLELVTQAIEACRKELSGTTTTT
SEQCEQTTSTVCGGPVFEEFIPIKKISSLCEEVQEEEEEDGEHESSPELVNNKSDWLRS
VQLWNHSPDLNPKBEERVAKKAKVVEVKPKSGAFQPFQKRVLETDLQPAVKVASSMPATTT
SSTTETCGGKSDLIKAGDEERRIEQQQSQSHTRKQRRCSPELHRRFLNALQQLGGSHV
ATPKQIRDHMKVDGLTNDEVKSHLQKYRLHTRRPAATSVAAQSTGNQQQPQFVVVGGIWW
PSSQDFPPPSDVANKGGVYAPVAVAQSPKRSLESCNSPAASSSTNTNTSTPVS*

>G2583 (38..607)

CAAAATCAGAAAATATAGAGTTTGAAGGAACTAAAAGATGGTACATTCGAGGAAGTTCCG
AGGTGTCCGCCAGCGACAATGGGGTTCTTGGGTCTCTGAGATTGCGCCATCCTCTATTGAA
GAGAAGAGTGTGGCTTGAAGCTTTGAAACGGCAGAAGCGGCTGCAAGAGCATACGACCA
AGCGGCTCTTCTAATGAACGGCCAAAACGCTAAGACCAATTTCCCTGTGCTAAAATCAGA
GGAAGGCTCCGATCACGTTAAAGATGTTAACTCTCCGTTGATGTCACCAAAGTCATTATC
TGAGCTTTTGAACGCTAAGCTAAGGAAGAGCTGCAAAGACCTAACGCCTTCTTTGACGTG
TCTCCGTCTTGATACTGACAGTTCCACATTTGGAGTTTGGCAGAAAACGGGCCGGGTGCGAA
ACAAGTCCGACTTGGGTCATGCGCCTCGAAGTTGGGAACGTAGTCAACGAAAGTGCGGT
TGACTTAGGGTTGACTACGATGAACAAACAAAACGTTGAGAAAGAAGAAGAAGAAGAAGA
AGCTATTATTAGTAGGATGAGGATCAGTTAGCTATGGAGATGATCGAGGAGTTGCTGAATTG
GAGTTGACTTTTGAAGCTTTTAACTTTGTTGCAAGTCCACAAGGGGTAAAGGTTTTTC

>G2583 Amino Acid Sequence (domain in AA coordinates:4-71)

MVHSRKFRGVRQRQWGSWVSEIRHPLLKRRVWLGTFTAEAAARAYDQAALLMNGQNAKT
NFPVVKSEEGSDHVKDVSPLMSPKSLSELLNAKLKRSCKDLTPSLTCLRLDTSSSHIGV
WQKRAGSKTSPTWVMRLELGNVNVESAVDLGLTTMNKQNEKEEEEEEAIIISDEDQLAME
MIEELNWS*

>G2724 (1..651)

ATGGAAATAGAAATAAGGAGAGGTCCATGGACTGTGGAAGAAGACATGAAGCTCGTCAGT
TACATTCTCTCTTCAAGGTGAAGGAAGATGGAAGCTCCCTCTCTCGTTCTGCTGGACTGAAT
AGAACGGGGAAAAGTTGCAGATTGCGGTGGCTAAATTATCTCCGGCCGGATATCCGCCGT
GGAGACATATCCCTTCAAGAACAAATTTATCATCCTTGAAGTCCATTCTCGTTGGGGAAAT
CGGTGGTCAAAGATTGCTCAACATTTACCGGGAAGAACAGATAACGAGATAAAGAATTAT
TGGAGAACACGTGTTCAAAAGCATGCAAACTTCTAAAATGTGACGTGAACAGCAAGCAA
TTCAAAGACACCATCAAACATCTCTGGATGCCTCGTCTCATCGAGAGAATCGCCGCCACT
CAAAGTGTCGAATTTACCTTAACCACTACTCGCCTGAGAACTCCAGCGTCGCCACCGCC
ACGTATCAACGTCGTCTGAGGCTGTGAGATCGAGTTTCTACGGTGGTGATCAGGTG
GAATTTGGAACGTTGGATCATATGACAAATGGTGGTTATTGGTTCAACGGCGGAGATACG
TTTGAAACTTTGTGTAGTTTTGACGAGCTCAACAAGTGGCTCATACAGTAG

>G2724 Amino Acid Sequence (conserved domain in AA coordinates:7-113)

MEIEIRRGPTVEEDMKLVSYISLHGEGRWNSLSRSAGLNRTGKSCRLRWLNLYLRPDIRR
GDISLQEQFIILELHSRWGNRWSKIAQHLPGRTDNEIKNYWRTRVQKHAKLLKCDVNSKO
FKDTIKHLWMPRLIERIAATQSVQFTSNHYPENSSVATATSSSTSSSEAVRSSFYGGDQV
EFGTLDHMTNGGYWFNGGDTFETLCSFDELNKWLIQ*

>G377 (1..396)

atgggtctctcgcatcttcccaacagcgtcagaaggagtactaccacttctggtgatgaac
acggttgtttcaatcactctgttgaagaacatggtaggtctgttttcaaatgttgca
tccgagactgaatcttccatggagatagacgacgagcctgaagatgattttgttactaga
agaatctcgataacacagttcaagtctctatgtgagaacatagaagaggaagaagaagag
aaaggtgtggagtgtgtgtgtgcctttgtgggtttaagaggaagaggaagtgtgtgag
ttggttcttgcgaagcatttcttccacagagctgtctagacaactggtttggttaataac

cacaccacatgccctctttgcaggtccattctctag

>G377 Amino Acid Sequence (domain in AA coordinates:85-128)
MGLSHFPTASEGVLP LLVMNTVVSITLLKNMVRSVFQIVASETESSMEIDDEPEDDFVTR
RISITQFKSLCENIEEEEEKGVCCVCLCGFKEEEEVSELVSKHFFHRACLDNWFGNN
HTTCPLCRSIL*

>G428 (97..1032)

TTACTTTTGTGTTTCTTCATATTCTTCAGAAGCAAGCACAAGGCTAGGGATCGAAGAAGC
GGCGATCACTGATCGTATCTCACTACGATCACATTAATGGATAGAATGTGTGGTTTCCGC
TCGACGGAAGACTATTTCGGAGAAAGCGACGTTGATGATGCCGTCCGATTATCAGTCTTTG
ATTTGTTCAACCACCGGAGACAATCAAAGACTGTTTGGATCCGACGAACCTCGCTACCGCT
TTGTCTCGGAGTTGCTTCCGCGTATTCGAAAAGCTGAGGATAATTTCTCTCTTAGTGTC
ATCAAATCCAAAATCGCTTCTCATCTCTTGTATCCTCGCTTACTCCAAACCTACATCGAT
TGCCAAAAGGTGGGAGCGCTATGGAATAGCGTGTATATTGGAAGAGATTGAGCGAGAG
AACCATGTGTACAAGAGAGATGTTGCTCCATTATCTTGCTTTGGAGCTGATCCTGAGCTT
GATGAATTCATGGAAACCTACTGTGATATATTGGTTAAATACAAAACCGATCTTGGGAGG
CCGTTTCGACGAGGCTACAACCTTTCATAAACAAGATTGAAATGCAGCTTCAGAACTTGTGC
ACTGGTCCAGCGTCTGCTACAGCTCTTTCAGATGATGGTGCGGTTTCATCTGACGAGGAA
CTGAGAGAAGATGATGACATAGCAGCGGATGACAGCCAACAAGAAGCAATGACCGCGAT
CTGAAGGACCAGCTACTACGCAAATTTGGTAGCCATATCAGTTTCATTGAAACTCGAGTTC
TCTAAAAAGAAGAAGAAAGGGAAGCTACCAAGAGAAGCAAGACAAGCGTTGCTCGATTGG
TGGAATGTTTCATAATAATAATGGCCTTACCCTACTGAAGGCGACAAAATAGCTCTGGCTGAA
GAAAACAGGTTTGGATCAAAAACAAATCAACAATTGGTTTATAAACCAAGGAAACGCCAT
TGGAAGCCTTCGGAGAACATGCCGTTTGTATGATGAGACGATTCTAATGAAACATTCTTT
ACCGAGGAATGAAAAGAGAGACATGGGATTGTGCTATTGTATAATTTTACACTGTTTTC
CAAGAAAAGAAAACAGTAAAAAGCTTTTGGTAAATGGGACATCATCGCAATGAATGGAA
CCAGTTAGCCAAAACGGTCAAGGGCGTGGCGTAACGAGACATTGTATTGGAATAGTGGC
AATATTATGTCACTAATCTTCCAATGGTCCAAAATGATAGATTCTTATTGTATTGAAC
CTTACTTAGATAGCTGATGTGTCACTAAATAATTTATTTTATCCTTATACTACTTGTA
TCAATGTCTCTAATTGATCAATTGTTGCTTGTCTATTCAAAAAAAAAAAAAAAAAAAAAA

>G428 Amino Acid Sequence (domain in AA coordinates: 229-292)
MDRMCGFRSTEDYSEKATLMPSDYQSLICSTTGDNRQLFGSDELATLSSSELLPRIRKA
EDNFSLSVIKSKIAHPLYPRLLQTYIDCQKVGAPMEIACILEEIQRENHVKRDVAPLS
CFGADPELDEFMETYCDIILVKYKTDLARPFDEATTFINKIEMQLQNLCTGPASATALSDD
GAVSSDEELREDDDDIAADDSQQRSDNRDLKDQLLRKFGSHISSLKLEFSSKKKKKGKLPRE
ARQALLDWNVHNWKPYPTEGDKIALAETGLDQKQINNWFINQRKRHWKPSNMPPFDM
DDSNETFFTEE*

>G447 (241..3501)

CTTTTTAAGAGCTTAAAAATTTGCTTTGAAGCTTCAAATATTCTTATGAACTAAAAAGAA
GAAAAAAGCTTTTGTCTTTTCTTCTTAGCAGCAGAAATGATTTTTGTTTCCAAAATTATT
ACTATTTAGTTTCTCTCGTGCTTCTCTTGAGCAAATACAGATTCTGTTAATTTTGCTGA
AGAAGAAGAACTCTGTTTCTTCCCTGCACCAAACCAATTTTTTCGTTCTTCTATAAACC
ATGAAAGCTCCATCAAATGGATTCTTCCAAGTTCCAACGAAGGAGAGAAGAAGCCAATC
AATTCTCAACTATGGCAGCTTGTGCAGGGCCTTTAGTTTCATTACCTCCTGTGGGAAGT
CTTGTGGTTTACTTCCCTCAAGGACACAGCGAGCAAGTTGCAGCATCGATGCAGAAGCAA
ACAGATTTTATACCAAATTACCCAAATCTTCTTCTAAGCTGATTGTGCTTGCTTCACAGT
GTTACATTACATGCTGATACCGAAACAGATGAAGTCTATGCACAAATGACTCTTCAACCT
GTGAATAAGTATGATAGAGAAGCATTGCTAGCTTCTGATATGGGCTTGAAGCTAAACAGA
CAACCTACTGAGTTTCTTGTGCAAGACTCTTACTGCAAGTGACACAAGCACTCATGGTGGA
TTCTCTGTACCGCGTCTGTCAGCTGAGAAAATATTCCTCCTCTTGATTCTCGATGCAA
CCGCTTGCAGAGAGATTGTAGCTAAAGATTACATGATACTACATGGACTTTCAGACAT
ATCTATCGAGGCCAACCAAAAAGACACTTGCTTACCACAGGTTGGAGCGTTTGTGTAGC
ACAAAGAGACTATTTCGGGTGATTTCAGTTTGTGTTGTAAGAGATGAGAAATCACAGCTG
ATGTTGGGTATAGACGTGCAAATAGACAAACTCCGACTCTTCTCATCGGTCATATCC
AGCGACAGTATGCACATTGGGATACCTGTCAGCTGCAGCTCATGCTAATGCCAATAGTAGC
CCTTTTACCATCTTCTTCAATCCAAGGGCAAGTCCTTCAGAGTTTGTAGTTCTTTAGCC
AAATACAACAAAGCCTTATACGCTCAAGTATCTCTAGGAATGAGATTCCGGATGATGTTT
GAGACTGAGGATTGTGGGGTTTCGTAGATATATGGGTACAGTCACAGGTATTAGTGATCTT

GACCCCTGTAAGATGGAAAGGCTACAATGGCGTAATCTTCAGGTAGGATGGGATGAATCA
ACACCTTGGAGATAGGCCAAGCCGAGTATCCATATGGGAAATCGAACCCGTCATAACTCCT
TTTTACATATGTCCTCCTCCATTTTTTCAGACCTAAGTACCCGAGGCAACCCGGGATGCCA
GATGATGAGTTAGACATGGAAAATGCTTTTCAAAAGAGCAATGCCTTGGATGGGAGAAGAC
TTTGGGATGAAGGACGCACAGAGTTCGATGTTCCCTGGTTTAAGTCTAGTTCAATGGATG
AGTATGCAGCAAAACAATCCATTGTTCAGGTTCTGCTACTCCTCAGTCCCGTCCGCGCTC
TCATCTTTTAACTTACCAACAATTTTGCTTTCCAACTGACCCCTTCCAAGCTGTTGAACTTC
CAATCCCCAAACCTCTCTTCCGCAAAATTTCCCAATTCAACAAACCGAACACGGTTAACCAT
ATCAGCCAAACAGATGCAAGCACAAACCAGCCATGGTGAAATCTCAACAACAACAACA
CAACAACAACAACAACAACAACAACAACATGCAACAACAACAACAACATCAGATG
TCACAGCAACAGGTGCAGCAACAAGGGATTTATAACAATGGTACGATTGCTGTTGCTAAC
CAAGTCTCTTGTCAAAGTCCAAACCAACCTACTGGATTCTCTCAGTCTCAGCTTCAGCAG
CAGTCAATGCTCCCTACTGGTGCTAAAAATGACACACCAAGAACATAAATCTATGGGAAT
AAAGGCTTGTCTCAAATGACATCGTTTTGCGCAAGAAATGCAGTTTCAGCAGCAACTGGAA
ATGCTAACAAGTAGCCAGCTATTATGAAGAACCGACAGAAACAGTCCCTCTCTCCATTCATTA
TGCAAAAATCTGTGCCAAAATCCTCAGCAACTCCAATGCAACAACAATCATCAAAACCA
AGTCCCTTCACAACAGCTTCAGTTGCAGCTACTGCAGAAGCTACAGCAGCAGCAACAGCAG
CAGTCGATTCTCTCAGTAAGCTCATCCTTACAGCCACAATTATCAGCGTTGCAGCAGACA
CAAAGCCATCAATTGCAACAACCTTCTGTCTGCTCAAATCAACAGCCCTTGGCACATGGT
AATAACAGCTTCCCAGCTTCAACTTTTCATGCAGCCTCCACAGATTCAGGTGAGTCCCTCAG
CAGCAAGGACAGATGAGTAAACAAAATCTGTAGCCGCTGGAAGTACATTTCTGGCCAC
ACAGATGGAGAGCTCCTTCTTGTTTCAACTCTCAGCTTCCGCCAAATAACACGGGACATGAT
ATAGTTTTCCCGCAAAATTTCTTGAGCAGAAATCAACAGCAAGGACAAGCTGCATCTGTA
TCTGCATCTGATTCACTCTTTGAGCGCGCAAGCAATCCGGTCCAAGAGCTTTATACAAAA
ACTGAGAGCCGGATCAGTCAAGGCATGATGAATATGAAGAGTGCTGGTGAACATTTTCAGA
TTTAAAGCGCGGTACAGATCAAATCGATGTATCCACAGCGGGAACGACGTACTGTCTCT
GATGTTGTTGGCCCTGTACAGCAGCAACAACTTTCCACTTCCATCATTTGTTTTGTAT
GGAGACTGCCAATCTCATCATCCAAGAACAACTTACGACTTTCCCTGTTAATCTCGAAGCC
GTAACTTCTGATCCACTTATTCTCAAAGGACTTTCAAACCTTGGTTCCCAACTATGGC
AACACACCAAGAGACATTGAGACGGAGCTGTCCAGTGCTGCAATCAGTTCTCAGTCATT
GGTATTCCCAGCATTTCCCTTTAAGCCCGGATGTTCAAATGAGGTTGGCGGCATCAATGAT
TCAGGAATCATGAATGGTGGAGGACTGTGGCCCAATCAGACTCAACGAATGCGAACATAT
ACAAAGGTTCAAAAACGAGGGTTCAGTAGGTAGATCAATAGATGTTACCCGTTATAGCGGC
TATGATGAACTTAGGCATGACTTAGCGAGAATGTTTGGCATCGAAGACAGCTCGAAGAT
CCGCTAACCTCTGATTTGGAACCTCGTCTACACCGCATCAGAAAACGATATTTTACTAGTT
GGTGATGATCCTTGGGAAGAGTTTGTGAAGTTCGCTGCAGAACATAAAGATACTATCATCA
GTAGAAGTTTCAGCAAATGAGCTTAGACGGAGATCTTGAGCTATCCCAACCACAAACCAA
GCCTGCAGCGAAACAGACAGCGGAAATGCTTGGAAAGTACACTATGAAGACACTTCTGCT
GCAGCTTCTTTCAACAGATAGAAAATAAAAGATGCAAAATATACCAAGTCAACTTACATTA
TCATTTCGAGGCCATCGCAAAGTACATGTTTTTTTTTGTGTGTATGTACTGCAAAACAA
ACTGAGAAGAAGAAGATATGACACGGTATATAAACATTTTATAGGACAGTGAATTTGATT
TTTCACTTCAACTTGATGTTGTTGTACTTTCTTGTTTCCATTTTGTATTTGATAACAGTATAA
TGCTTGCAAGTCTATGAGGACATATCTTATACAGAGATACTAAGATGTAATGTTAATG
TAACTAAACAATTACCTTCATTAATCATGAATCCTTTGGTTCGTTTAAAA

>G447 Amino Acid Sequence (conserved domain in AA coordinates:22-356)

[illegible]

QQGQMSNKNLVAAGRSHSGHTDGEAPSCSTSPSANNTGHDNVSPTNFLSRNQQQGQAASV
SASDSVFERASNVPQELYTKTESRISQGMNMKSAGEHFRFKSAVTDQIDVSTAGTTYCP
DVVGVPVQQQTFPLPSFGFDGDCQSHHPRNNLAFPGNLEAVTSDPLYSQKDFQNLVPNYG
NTPRDIETELSSAAISSQSFGPSIPFKPGCSNEVGGINDSGIMNGGGLWPNQTQRMRTY
TKVQKRGSVGRSIDVTRYSGYDELRLHDLARMFGIEGQLEDPLTSDWKLVTYTDHENDILLV
GDDPWEEFVNCVQNIKILSSVEVQMSLDGDLAAIPTTNQACSETDSGNAWKVHYEDTSA
AASFNR*

>G464 (41..760)

CTCTGCTGGTATCATTGGAGTCTAGGGTTTTGTTATTGACATGCGTGGTGTGTGAGAATT
GGAGGTGGGGAAGAGTAATCTTCCGGCGGAGAGTGAGCTGGAATTGGGATTAGGGCTCAG
CCTCGGTGGTGGCGCGTGGAAAGAGCGTGGGAGGATTCTTACTGCTAAGGATTTTCCTTC
CGTTGGGTCTAAACGCTCTGCTGAATCTTCTCTACCAAGGAGCTTCTCCTCCTCGTTC
AAGTCAAGTGGTAGGATGGCCACCAATTGGGTTACACAGGATGAACAGTTTGGTTAATAA
CCAAGCTATGAAGGCAGCAAGAGCGGAAGAAGGAGACGGGGAGAAGAAAGTTGTGAAGAA
TGATGAGCTCAAAGATGTGTCAATGAAGGTGAATCCGAAAGTTCAGGGCTTAGGGTTTGT
TAAGGTGAATATGGATGGAGTTGGTATAGGCAGAAAAGTGGATATGAGAGCTCATTCTGTC
TTACGAAAACCTTGGCTCAGACGCTTGAGGAAATGTTCTTTGGAATGACAGGTACTACTTG
TCGAGAAAAGGTTAAACCTTTAAGGCTTTTAGATGGATCATCAGACTTTGTACTCACTTA
TGAAGATAAGGAAGGGGATTGGATGCTTGTGGAGATGTTCCATGGAGAATGTTTATCAA
CTCGGTGAAAAGGCTTCGGATCATGGGAACCTCAGAAGCTAGTGGACTAGCTCCAAGACG
TCAAGAGCAGAAGGATAGACAAAGAAACAACCCTGTTAGCTTCCCTTCCAAAGCTGGCA
TTGTTTATGTATTGTTTGGAGTTTGCAATTTACTCGATACTTTTGAAGAAAGTATTTTG
GAGAATATGGATAAAAGCATGCAGAAGCTTAGATATGATTGAATCCGGTTTTCGGATAT
GGTTTTGCTTAGGTCATTCAATTCGTAGTTTTCCAGTTTGTTCCTTTGGCTGTGTAC
CAATTATCTATGTTCTGTGAGAGAAAGCTCTT

>G464 Amino Acid Sequence (domain in AA coordinates: 20-28, 71-82, 126-142, 187-224)

MRGVSELEVKGKSNLPAESELELGLGLSLGGGAWKERGRILTAKDFPSVGSKRSAESSHQ
GASPPRSSQVVGWPPIGLHRMNSLVNNQAMKAARAEEGDGEKKVVKNDLKDVSMMKVNPK
VQGLGFVKVNMNDGVGIGRKVDMRAHSSYENLAQTLEEMFFGMTGTTCREKVKPLRLLDGS
SDFVLTYEDKEGDWMLVGDPWRMFINSVKRLRIMGTSEASGLAPRRQEQKDRQRNPNV*

>G557 (192..698)

CAGAGATCTGACGGCGGTAGCAGAGTAATCTATTCTTCCCAAATGTCTCGCAATTAGA
TTCTTTCCAAAGTTCTTCTGTAAATCCCAAGTCCCGCTCTTTCTCTTTATCCTTTTCAC
CAGCTTCGCTACTAAGACAACAAATCTTCCCTCTCTCTCTCGCCTGATCGATCTTCAA
GAGTAAGAAAAATGCAGGAACAAGCGACTAGCTCTTTAGCTGCAAGCTCTTACCATCAA
GCAGCGAGAGGTATCAAGCTCTGCTCCACATTTGGAGATCAAAGAAGGAATTGAAAGCG
ATGAGGAGATACGGCGGAGTGCCGGAGTTTGGAGGAGAAGCTGTGCGGAAAAGAACTTCCG
GTAGAGAATCTGGATCGGCGACCGGTGAGGAGCGGACACAGGCGACTGTGCGAGAAAGTC
AAAGGAAGCGAGGGAGGACACCGCGGAGAAAGAGAACAGCGGCTGAAGAGGTGTTGA
GGAACAGAGTTTCAGCTCAGCAAGCAAGAGAGAGGAAAAGGCTTACTTGAGCGAGTTGG
AAAACAGAGTGAAGAGCTTGGAGAACAAAACTCTGAACCTGAAGAGCGACTCTCTACTC
TTCAGAACGAGAACAGATGCTTAGACATATTCTGAAGAACACAACAGGAAACAAGAGAG
GAGGTGGTGGTGGTTCTAATGCTGATGCAAGCCTTTGATCTCCTTCTTCTTGTGTTA
TATTTTGTGGATAAAATTTACAGAGAATTGTATCAATAATTATCATGTTAAAATTATAT
GGGATGTGAGAGCTAATATTGCAATTGTAGACCAAGTTCTTTAAAAA

AAAAA

>G557 Amino Acid Sequence (domain in AA coordinates: 90-150)
MQEQATSSLAASSLPSSSSERSSSSAPHLEIKBGIESDEEIRRVPFEGGEAVGKETSGRES
GSATGQERTQATVGESQKRGRTPAEKENKRLKRLLRNVSAQQARERKKAYLSELENRV
KDLENKENSELEERLSTLQENQMLRHILKNTTGNKRGGGGGSNADASL*

>G577 (44..2155)

AAAAACAGACTGAGAGAGAGAGAGAGAGTGTGTTGTTGGCCATGGGATGCACGGCCTC
CAAGCTCGACAGTGAGGATGCTGTCCGTGCAAGGAGCGGCGCGCTTTATGAAGGA
CGCCGTCTACGCTCGTCAACATCTCGCCCGCGCTCACTCTGACTACTGCCGCTCCCTTCG
TCTCACTGGCTCTGCCCTCTCCTCCTTCGCCCGCGGAGCCCTCTCCGTCTCCGAGAA
TACTCCGCTGTTTTCTCCGCCCTTCTCCAGTCAGGACGCGCCACGTGTCCCTTCTTC

CCATTCCCCAGAACCCCTCCTCCGCCATCCGCAGCAAGCCTAAGCCTACTAGGCCTAG
GAGGCTTCCACACATTCTCTCCGACTCCTCTCCTTCTTCTCCTGCCACCAGTTTCTA
TCCCCTGCTCACCAGAAGCTTACTTCTCGCTCTCCATCTCAAGCTTCTCTGTCTG
GAATGGGAGAATTTCTACCTCCCTCTCCCCCGACTCCGAGTACTTCGAACGCAAAGC
TCGCCAGAACCAAGCACCGTCTCCTTCCGACTACGACGCCGAAAGTGAAGATCCGA
CCACGATTACTGCCACTCACGGAGAGATGCCGCCGAGGAAGTTCACTGCAGCGAGTGGGG
CGACGACCACGACCGTTTCACTGCCACCTCTTCGTCCGACGGAGATGGGGAGGTGAAAC
TCACGTTTCCAGATCCGGTATTGAAGAAGAGCCTGTGAAACAACCACATCAAGACCCAAA
TGGCAAAGAGCACTCTGACCATGTTACCACTTCTTCCGACTGCTACAAGACCAAATTGGT
GGTAAGGCACAAGAATTTGAAGGAGATCCTTGACGCCGTTCAAGACTACTTCGACAAGGC
TGCTTCCGCTGGGGACCAAGTCTCCGCCATGCTTGAGATCGGCCGGGCTGAGCTCGACCG
CAGCTTCAGCAAGCTGAGGAAGACGGTGTATCATTCAAGCAGTGTGTTTCAAGCACTTGAG
CGCAAGCTGGACCTCAAACCCCCATTGGCAGTCAAATACAAGCTCGATGCATCTACCTT
GAATGATGAACAAGGCCGCTCAAGAGCCTCTGCTCCACTCTAGACCGACTCCTCGCTTG
GGAAAAGAAGCTTTATGAGGATGTCAAGGCAAGAGAAGGAGTTAAGATTGAGCAGAGAA
GAAGCTGTCTGCGTGCAGAGTCAAGGATATAAGGGAGGTGATGAATCCAAGCTAGACAA
GACTAAAACCTTCATAACCAGACTGCAATCACTCATCATTGTTTCTTCAGAAGCTGTTT
AACCACGTCTAATGCCATTCTCCGCCTCCGGGACACTGACCTTGTCCTCAGCTTGTTGA
ACTCTGCCACGGATTAATGTACATGTGGAAGTCAATGCACGAGTATCACGAAATCCAGAA
CAACATCGTGCAACAAGTCCGTGGCTGATCAACCAAACAGAGAGAGGTGAGTCAACATC
AGAGGTACACCGGCAGGTGACGCGGACCTAGAGTCAAGTGTGTCTTGTGGCATTTCGAG
CTTCTGTGCGATCATTAAATTCAGAGGGAGTTTCAATATGCTCTCTCCACGCATGGTTCAA
GCTGAGCCTGGTTCCCTGAGCAACGGAGACCCAAAGAAACAGCGGCCAGACTCATTTCG
CTTGTGCGAGGAGTGAAGCAGAGCCTGGAACGGGTGCTGACACAGTGGCGTCAAGAGC
CATAAAGAGCTTTGTAAACGTGGTACATGTGATATCAATAAAGCAGGCGGAAGAGGTGAA
GATGAAGAAACGCACGGAGAGTGCAGGAAAGGAGCTGGAGAAGAAAGCATCCTCACTGAG
GAGCATAGAGAGGAAGTACTACCAGGCATACTCGACGGTTGGGATAGGCCCTGGACCGGA
GGTGTGAGTCAACGGGACCCGCTATCTGAGAAGAAATGTGAGCTGGCGGCATGTCAAG
GCAGGTGAGGATGAGGTAATGAGGCACGTGAAGGCTGTGGAGGTGACACGAGCTATGAC
TCTCAACAATCTACAAACCGCCTGCCAATGTATTCCAGGCCTTGACCAGCTTCTCATC
TCTCTTCACTGAATCTCTCCAGACTGTCTGTTCTCGTTCTACTCCATCAACTGATTATG
TCCAAGTTTCTCATTTATTTTAAGCTCTCATTACGTTGGTATCATGTAAATTTGAGGAT
TGATTAAATTGAGTCTTGTGGTTTTGTGAGGACTCACAATCTTTCTCATTTAAAAA
AAAAA

>G577 Amino Acid Sequence (domain in AA coordinates: TBD)

MGCTASKLDSSEDAVRRCKERRRLMKDAVYARHHLAAHSDYCRSLRLTGSALSSFAAGEP
LSVSENTPAVFLNRPSSQDAPRVPSHSPPEPPPIRSKPKPTRPRRLPHILSDSSPSSS
PATSFYPTAHQNSYSSRSPQASSVWNWENFYPPSPDSEYFERKARQNHKRPSPSYDA
ETERSDHDYCHSRRDAEEVHCSEWGDDHDFRFTATSSSDGDGEVETHVSRSGIEEFPVKQ
PHQDPNGKEHSDHVTSSDCYKTKLVVRHKNLKEILDVQDYFDKAASAGDQVSAMLEIG
RAELDRSFSKLRKTVYHSSSVFNLASWTSKPPLAVKYKLDASTLNDEQGGGLKSLCSTL
DRLLAWKKLYEDVKAREGVKIEHEKKLSALQSQEYKGGDESKLDTKTSITRLQSLIIV
SSEAVLTSSNAILRLRDLVLPQLVELCHGLMYMWMKSMHEYHEIQNNIVQQVRGLINQTE
RGESTSEVHRQVTRDLESVSLWHSSFCRIIKFQREFICSLHAWFKLSLVPLSNGDPKKQ
RPDSFALCEWQSLERVPDVTVAEAIKSFVNVVHVISIKQAEVVKMKRTESAGKELEK
KASSLRSIERKYYQAYSTVIGPGPEVLDSRDPLSEKKCELAACQRQVEDEVMRHVKA
VTRAMTLNQLTGLPNVFQALTSFSSLFTESLQTVCSRSYSIN*

>G674 (1..786)–

ATGGTGTTTAAATCAGAAAAATCAAACCGGGAATGAAATCAAAGGAGAAGCAAAGGAAG
GGATTATGGTCACCCGAGGAAGATGAGAAGCTTAGGAGTCATGTCTCAAATATGGCCAT
GGATGCTGGAGTACTATTCTCTTCAAGCTGGATTGCAGAGGAATGGGAAGAGTTGTAGA
TTAAGGTGGGTAAATTATTTAAGACCTGGACTTAAGAAGTCTTTATTCACTAAACAAGAG
GAACTATACTTCTTCACTTCATTCCATGTTGGGTAAACAATGGTCTCAGATATCGAAA
TTCTTACCAGGAAGAACCAGACAGAGATCAAAAACCTATTGGCATTCTAATCTAAAGAAG
GGTGTAACTTTGAAACAACATGAAACCACAAAAAACATCAAAACACCTTTAATCACAAAC
TCACTTGAGGCCTTGACAGAGTTCAACTGAAAGATCTTCTTCATCTATCAATGTGCGAGAA
ACGTCTAATGCTCAAACCTCAAGCTTTTCGCCAAATCTCGTGTCTCGGAATGGTTAGAT

CATAGTTTGCTTATGGATCAGTCACCTCAAAAGTCTAGCTATGTTCAAAATCTTGTTTTA
CCGGAAGAGAGAGGATTCAATTGGACCATGTGGCCCTCGTTATTTGGGAAACGACTCTTTG
CCTGATTTTCGTGCCAAATTCAGAATTTTGTGGATGATGAGATATCATCTGAGATCGAG
TTCTGTACTTCATTTTCAGACAACTTTTGTTCGATGGTCTCATCAACGAGCTACGACCA
ATGTAA

>G674 Amino Acid Sequence (domain in AA coordinates: 20-120)
MVFKEKSNREMKSKEKQKRLWSPEEDEKLRSVHLKYGHGCWSTIPLQAGLQNRNGKSCR
LRWVNYLRPGLKKSLFTKQEEITILLSLHSM LGNKWSQISKFLPGRTDNEIKNYWHSNLKK
GVTLKQHETTKKHQTPLITNSLEALQSSTERSSSSINVGETSNAQTSSFSFNLVFEWLD
HSLMDQSPQKSSYVQNLVLPBERGFIGPCGPRLGNDSLPDFVPNSEFLDDEISSEIE
FCTSFSDNFLFDGLINELRPM*

>G736 (1..513)
ATGGCGACTCAAGATTCTCAAGGATTAAACTCTTTGGCAAACTATTGCATTTAACACT
CGAACAAATAAAAAATGAAGAAGAGACACACCCGCCGAGCAAGAAGCCACAATAGCCGTT
AGATCATCATCATCATCGGATCTGACGGCCGAGAAGCGTCCGGATAAGATCATAGCATGT
CCAAGATGCAAGAGCATGGAGACAAAGTTCTGTTACTTCAACAAC TACAACGGTAATCAG
CCTCGACACTTTTGTAAAGGCTGCCACCGTTACTGGACCGCCGGTGGTGCCTCCGGAAC
GTTCCCGTCCGGCGCCGGTCTGTCGGAAGTCAAACCACTGGTCTGTCGTGGTGGTATG
CTTGGAGATGGAAATGGTGTTCGCCAAGTCGAGCTTATAAATGGCTTGTCTCGTTGAGGAG
TGGCAGCATGCCGACCCGACGTCACGGTAGTTTCCGGCATGATTTTCCCATGAAGCGG
CTCCGGTGTACTCCGACGGTCAATCGTGCTGA

>G736 Amino Acid Sequence (domain in AA coordinates: 54-111)
MATQDSQGIKLPFKTIAFNTRTIKNEETHPPEQEIIVRSSSSDLTAEKRPDKIIAC
PRCKSMETKFCYFNNGNQP RHFCKGCHRYWTAGGALRNVPGAGRRKSKPPGRVVGVM
LGDGNGVRQVELINGLLVEEWQHAAAAHGSFRHDFPMKRLRCYSDGQSC*

>G903 (96..1496)
CCCGGGTCGACCCACGCGTCCGCTCTCTCTCTGAACTATACAAAAACCTACTTTTAAT
TTCTCTTCCAAGAAGTCAAGAACCCAGAAGAAGACATGACAAGTGAAGTTCTTCAAAACA
TCTCAAGTGGATCAGGTTTTGTCTAGCCACAGAGCTCATCAACCCTGGATCATGATGAAT
CTCTCATCAATCCTCTCTTGTAAAGAAAAGAGAAATCTCCCTGGAAATCCTGATCCGG
AAGCTGAAGTGATAGCTTTATCCCCACGACCTTGATGGCTACGAACCGGTTCTATGTG
AGGTATGTGGCAAAGGTTTCAAAGAGACCAAACTTACAGCTTCATCGGCGAGGACATA
ATCTTCCATGGAGTTGAAGCAGAGGACAAGCAAAGAGTGAGAAAACGTGTCTACGTTT
GCCCCGAGAAGACATGTGTCCACCATCACTCCTCTAGAGCTCTAGGCGATCTCACTGGAA
TCAAAAAGCATTTTTCGCCGAAACACGGGGAGAAGAAGTGACGTGCGAGAAATGTGCTA
AGAGATACGCAGTCCAATCTGATTGGAAAGCTCATTCCAAGACTTGTGGTACTAGAGAGT
ACCGTTGCGATTGTGGCACCATTCTTCAAGGCGAGACAGCTTTATCACTCATAGAGCTT
TCTGCGATGCCTTAGCGGAAGAAACCGCTAAGATAAACGCAGTGTCTCATCTCAACGGTT
TAGCCGCGGCTGGAGCCCCAGGATCAGTTAATCTCAACTATCAATATCTCATGGGAACAT
TCATCCCACCGCTTCAACCATTGTACCAACAACCGCAAACAAATCCAAACCATCATCATC
AACATTTTCAGCCACCAACTTCTTCGTCTCTCTCTATGGATGGGACAAGATATCGCGC
CGCCTCAACCGCAACCGGACTACGATTGGGTTTTTGGAAACGCTAAGGCAGCGTCTGCTT
GCATTGATAATAATAATACTCACGATGAGCAGATTACGCAAAACGCAACGCAAGTTTGA
CCACTACCCTACTCTCTCTGCCCCCTCTTTATTACGACGCGACCAACCAACCAACGCAA
ACGCAAATTCAAACGTGAATATGTCCGCGACAGCTTTACTACAGAAAGCTGCTGAAATTG
GCGCTACTTCTACAACAACCGCAGCGACCAATGACCCATCAACGTTTCTTCAAAGTTTCC
CGCTTAAATCCACCGATCAAACCACAGTTATGACAGTGGCGAAAAGTTTTTGTCTTGT
TCGGGTCTAACAACAACATTGGGTTAATGAGTCGTAGTCATGATCATCAAGAGATCGAGA
ACGCTAGAAATGACGTTACGTTGCGTCTGCCTTGGATGAATTACAGAATTACCCCTTGG
AACGTAGAAGAGTTGATGGTGGAGGTGAAGTGGGTGGAGGAGGGCAAACCTCGGGATTTC
TCGGGGTTGGTGTACAAACGTTGTGCCATCCATCGTCTATCAATGGATGGATTGAAAGA
GTTTAAAAATTTCCGGGTTAATGCATAAATTACGTAAAAGAAGAAGGAATCTTTTGTGAT
TTCCACCATTTTCTAAGATAACATATGTATATGGTAATGGAAGTTGTTTTCTTTTATTA
TTCAATATTCTAAACTTATGATATATGTATAATGAATGTGTTTATCTTCAA

>G903 Amino Acid Sequence (domain in AA coordinates: 68-92)
MTSEVLQTISSGSGFAQPQSSSLDHDHESLINPPLVKKRNLPGNPDPEAEVIALSPPTL
MATNRFLCEVCGKGFQDQNLQLHRRGHNL PWKLKQRTSKEVRKRVYVCPEKTCVHHSS

RALGDLTGIIKKHFCRKHGEKKWTCEKCAKRYAVQSDWKAHSKTCGTREYRCDCGTIFSRR
DSFIITHRAFCDALAEETAKINAVSHLNGLAAGAPGSVNLNYQYLMGTFIPLQPFVPPQ
QTNPNNHHQHFPPTSSSLSLWMGQDIAPPQPDYDWFVGNAAKASACIDNNNTHDEQI
TQNANASLT'TTTLSAPSLFSSDQFQONANANSNVNMSATALLQKAAEIGATSTTTAATND
PSTFLOQSFPLKSTDQTTSDGKFFALFGSNNNIGLMSRSHDHQEIENARNDVDVASAL
DELQNYPWKRRRVDGGGEGQTRDFLGVGVQTLCHPSSINGWI*

>G917 (32..679)

TTAGGGTTT'TAGAAAGATAGATCGATTGAAGATGAGGAAAGGTAAGAGAGTGATAAAAAA
GATAGAGGAGAAAAATAAGAGACAAGTGACATTTCGCAAAGAGAAAGAGAGTCTAATCAA
GAAGGCATATGAACCTCTCTGTTCTCTGCGATGTCCACCTGGTCTCATCATCTTCTCTCA
CTCCAACAGGCTCTACGATTTCTGCTCCAACCTCTACCAGCATGGAGAATCTCATCATGAG
ATACCAAAAGGAAAAAGAAGGTCAAACCACTGCAGAACACAGTTTCCACTCGGATCAGTG
TTCAGATTGCGTGAAGACGAAGGAATCAATGATGAGAGAGATAGAGAATCTTAAGCTGAA
TCTTCAATTGTACGACGGACATGGCTTGAATCTCTTGACCTACGACGAGCTCCTTTCTTT
TGAGCTCCATCTCGAATCTTCTCTACAACATGCTCGAGCTCGCAAGTCTGAGTTCATGCA
TCAGCAGCAGCAGCAACAAACAGATCAAAGCTTAAGGGAAGAAAGGGTCAAGGAAG
CTCTTGGGAGCAGCTGATGTGGCAAGCAGAGAGACAGATGATGACGTGTCAAAGACAAAA
AGATCCTGCGCCGCGAATGAAGGAGGAGTTCTTTTTTACGGTGGGGAACAACCCACCG
ACGTTCTTTCACCTCCTTAAGCTACCACAACCCAGGCCCAATACAGGCCCATAACTTCTCT
CTATCTATAAAAAACAACCTGATAGTAAAAAGTATTGACCCGGTTTGGTTTCGGTTATGTTG
ATACCAGACTATTAATTAACCTCGGTTAGACGTATTTACGACTTGATGCTATCTAGACCT
TTTTGCCCTTCAAAAAA

>G917 Amino Acid Sequence (conserved domain in AA coordinates:2-57)

MRKGKRVIKKIEEKIKRQVTFKRKSLIKKAYELSVLCDVHLGLIIFSHSNRLYDFCSN
STSMENLIMRYQKEKEGQTTAEHSFHSQSDCVKTKESMMREIENLKLNLQLYDGHGLN
LLTYDELLSFELHLESSLQHARARKSEFMHQQQQQQTDQKLKGKEKGQSSWEQLMWQAE
RQMMTCQRQKDPAPANEQGVFPLRWGTTTHRRSSPP*

>G921 (116..1024)

CCAAGATCGACTCTTACTTGAATCTCTCTCAACTTTCTTCTCAGCTTACGGGAACCTTC
CACACATATACATCCACAAGAACCATATCGAAGATTCATCCTACATATATTTACATGGA
TCAGTACTCATCTCTTTGGTTCGATACTTCATTAGATCTCACTATTGGCGTTACTCGTAT
GCGAGTTGAAGAAGATCCACCGACAAGTGCTTTGGTGAAGAATTAACCGAGTTAGTGC
TGAGAACAAGAAGCTCTCGGAGATGCTAATTTGATGTGTGACAACTACAACGCTTTGAG
GAAGCAACTTATGGAATATGTTAACAAGAGCAACATAACCGAGAGGGATCAAATCAGCCC
TCCCAAGAAACGCAATCCCCGGCGAGAGAGGACGCATTACGCTGCGCGGTTATTGGCGG
AGTGTCGGAGAGTAGCTCAACGGATCAAGATGAGTATTTGTGTAAGAAGCAGAGAGAAGA
GACTGTTCGTGAAGGAGAAAGTCTCAAGGGTCTATTACAAGACCGAAGCTTCTGACACTAC
CCTCGTTGTGAAGATGGGTATCAATGGAGGAAATATGGACAGAAAGTGAAGTAGAGACAA
TCCATCTCCAAGAGCTTACTTCAAATGTGCTTGTGCTCCAAGCTGTTCTGTCAAAAAGAA
GGTTCAGAGAAGTGTGGAGGATCAGTCCGTGTTAGTTGCAACTTATGAGGGTGAACACAA
CCATCCAATGCCATCGCAGATCGATTCAAACAATGGCTTAAACCGCCACATCTCTCATGG
TGGTTCAGCTTCAACACCCGTTGCAGCAACAGAGAAGTAGCTTGACTGTGCCGGTGAC
TACCGTAGATATGATTGAATCGAAGAAAGTGACGAGCCCAACGTCAAGAATCGATTTTCC
CCAAGTTCAGAACTTTTGGTGGAGCAAATGGCTTCTTCTTAACCAAAGATCCTAACTT
TACAGCAGCTTTAGCAGCAGCTGTTACCGGAAAATGTATCAACAGAATCATACCGAGAA
ATAGTTTAGCTTCAAATTCGGTTAGAGTTTTTAGATTTGAATTTGTATGAGTAAGAGAA
AGAGAGTAGATTATAATCCNTTGTGATACTGAAAAAAAAAAAAAAAAAAAAA

>G921 Amino Acid Sequence (domain in AA coordinates: 146-203)

MDQYSSSLVDTSLDLTIGVTRMRVEEDPPTSALVEELNRVSAENKKLSEMLTLMCDNINV
LRKQLMEYVKNKSNITERDQISPPKKRKS PAREDAFSCAVIGGVSESSSTDQDEYLCKKQR
EETVVKEKVSRYVYKTEASDTTLVVKDGYQWRKYGQKVTNRDNPSPRAYFKCACAPSCSVK
KKVQRSVEDQSVLVATYEGEHNHPMPSQIDSNNGLNRHISHGGSASTPVAANRRSSLTVP
VTTVDMIESKKVTSPTSRIIDFPQVQKLLVEQMASSLTDPNFTAALAAAVTGKLYQQNHT
EK*

>G922 (1..1449)

ATGGTGGCTATGTTTCAAGAAGATAATGGAACATCTTCTGTAGCTTCATCACCCTTCAA
GTCTTCTCAACTATGTCACTCAACAGACCGACTCTCCTCGCTTCTTCATCTCCGTTTCAT

TGTCCTCAAAGATCTCAAACCAGAGGAGCGTGGTCTCTACTTAATCCACCTCTTGCTAACT
TGTGCCAACCCAGTGGCTTCAGGTAGCCTCCAAAACGCTAACGCAGCGCTCGAGCAGCTC
TCTCACCTCGCTTCTCCTGACGGCGACACGATGCAGCGAATCGCTGCTTACTTCACCGAA
GCGCTTGCTAACAGAATCCTTAAGTCTTGGCCTGGTCTTTACAAGGCTCTTAACGCAACT
CAGACAAGAACTAACAAATGTCTCTGAGGAGATTCTGTTAGAAAGACTCTTCTTTGAGATG
TTCCCGATACTCAAAGTCTCTTACTTGCTCACTAATCGAGCTATACTCGAGGCTATGGAA
GGAGAGAAGATGGTTCATGTGATTGATCTCGATGCTTCTGAGCCAGCTCAATGGCTTGCT
TTGCTTCAAGCTTTTAACTCTAGGCCTGAAGGTCCACCTCATTTGAGAATCACTGGTGT
CATCACCAGAAGGAAGTGCTTGAACAAATGGCTCATAGACTCATTGAGGAAGCAGAGAAA
CTCGATATCCCGTTTCAGTTTAAATCCCGTTGTGAGTAGGTAGACTGTTTAAATGTAGAA
CAGTTGCGGGTTAAAACAGGAGAGGCCTTAGCCGTTAGCTCGGTTCTTCAATGCATACC
TTCTTGGCCTCTGATGATGATCTCATGAGAAAGAACTGCGCTTTACGGTTTCAGAACAAAC
CCTAGTGGAGTTGACTTGCAGAGAGTTCTAATGATGAGCCATGGCTCTGCAGCTGAGGCA
CGTGAGAATGATATGAGTAACAACAATGGGTATAGCCCTAGCGGTGACTCGGCCTCATCT
TTGCCTTTACCAAGTTCAGGAAGGACTGATAGCTTCTCAATGCTATTTGGGGTTTGTCT
CCAAAGGTCTGTTGGTCACTGAGCAAGACTCAGACCACAACGGCTCCACACTAATGGAG
AGGCTATTAGAATCACTTTACACCTACGCAGCATTGTTTGTGCTTGGAAACAAAAGTT
CCAAGAACGTCTCAAGATAGGATCAAAGTGGAGAAGATGCTCTTCGGGGAGGAGATCAAG
AACATCATATCTTGCAGGGGATTTGAGAGAAGAGAAAGACACGAGAAGCTTGAGAAATGG
AGCCAGAGGATCGATTTGGCTGGTTTTGGAATGTTCTCTTAGCTATTATGCGATGTTG
CAGGCTAGGAGATTGCTTCAAGGGTGGGTTTGTATGGGTATAGAATCAAGGAAGAGAGC
GGGTGCGCAGTAATTTGCTGGCAAGATCGACCTCTATACTCGGTATCAGCTTGGAGATGC
AGGAAGTGA

>G922 Amino Acid Sequence (conserved domain in AA coordinates:225-242)

MVAMFQEDNGTSSVASSPLQVFSTMSLNRPTLLASSSPFHCLKDLKPEERGLYLIHLLLT
CANHVASGSLQANANAALQLSHLASPDGDMQRIAAFYTEALANRILKSWPGLYKALNAT
QTRTNVSEEHVRRLLFFEMFPILKVSYLLTNRAILEAMEGEKMHVHIDLDASEPQWLA
LLQAFNSRPEGPPHLRITGVHHQKEVLEQMAHRLIEEAELDIPFQFNPVVSRLDCLNVE
QLRVKTGEALAVSSVLQLHFTLASDDDLMRKNCALRFQNNPSGVLDLQRVLMMSHGSAAEA
RENDMSNNNGYSPGDSASSLPLSSGRDTSFLNAIWGLSPKVMVTEQSDHNGSTLME
RLLESLYTYAALFDCLETQVPRTSQDRIKVEKMLFGEEIKNIISCEGFERRERHEKLEKW
SQRIDLAGFGNVPLSYAMQLQARRLLQCGCFDGYRIKEESGCAVICWQDRPLYSVSAWRC
RK*

>G932 (206..1213)

CCACGCGTCCGACCCTTGTACCTCTTTGTCTTAAGTACTCTTTAACCTACAATTCCT
AAGCTCTCAAGCCACAAAAACCACAAACCGTTCTTCACCAATATATATCTGATCATC
ATCAAAGTCTTCTCTCTGCTCATACCACAAACCGTTCCATTCTTCCCTAATCACAAG
TGATATTTACATAGAGAAGATAGAGATGGGAAGACCACCATGCTGTGACAAGATTGGAGT
GAAGAAAGGACCATGGACACCAGAGGAAGATATCATCTTGGTTTCTTACATCCAAGACA
TGGTCTCTGAAACTGGAGATCTGTGCCTACTCACACAGGTTTGGAGAGATGTAGCAAAAG
CTGTAGATTGAGGTGGACTAATTATCTTCCGACCTGGGATCAAGCGTGGAATTTACCCGA
GCATGAAGAGAAGATGATTCTCCATCTTCAAGCTCTTTGGGAAACAGGTGGGCAGCTAT
AGCATCATATCTTCCAGAAAGGACAGACAATGATATAAAGAACTATTGGAACACTCATTT
GAAGAAAAAGCTCAAGAAGATGAATGATTCTTGTGATAGTACTATCAACAATGGCCTTGA
TAATAAAGACTTCTCCATATCAACAAAAACACTACCTCACATCAAAGCAGCAACTCCAG
TAAAGGTCAATGGGAGAGAAGGCTTCAGACAGATATCAACATGGCTAAACAAGCTCTTTG
TGATGCCTTGTCTATTGACAAACCACAAAACCACTAATTTTCTATTCCCGATCTTGG
TTATGGTCCATCATETTTCTCGTCTCTACCACCACCACCACCACCACCACCACGAG
AAACACTAATCCATACCCATCTGGGGTCTATGCTTCAAGTGCTGAGAACATTGCTCGTTT
GCTTCAGAAATTTATGAAAGACACACCAAGACCTCGGTGCCCTTGCCGGTTGCGCCAC
CGAGATGGCTATCACCACGGCAGCTTCGAGCCCTAGCACAACCGAAGGAGACGGAGAAGG
GATTGACCATTCTTTGTTTCAGCTTCAACTCCATAGATGAAGCTGAAGAGAAGCCTAAACT
AATAGACCATGACATTAATGGTCTAATTACACAAGGCTCTCTTTCTTTGTTTCGAGAAATG
GCTCTTTGATGAGCAAAAGCCACGATATGATCATCAATAACATGTCACTAGAGGGTCAGGA
AGTGTGTGTTCTAGAAAGCATTAAGTTTGACGATTTGCTTGAGGAACCACGAGGCTTAGT
TATAAACAATTTGTATAATTAAGTACTCTTTAGTTTTGTTTTCAATCCTTATTATGATCA
TATTGCAGTAATTAGGGATTTTAGTCTTTAGTAGTAACCTTAAGTTTAAACATTTTTT

CTCTATCTTTTCTAGTAGTAACTCTTTATTTTCTTAAATCTTTGTGACGTGGAGATG
ATATCTTCTATGTAGTAGAACTCAAAGTGATCATCTTTATTAATGTAACGTCTTT
TTAAAAAAAAAAAAAAAAAAAA

>G932 Amino Acid Sequence (domain in AA coordinates: 12-118)
MGRPPCCDKIGVKKGPWTPPEEDIILVSYIQEHGPGNWRVPTHTGLRRCSKSCRLRWNTY
LRPGIKRGNFTEHEEKMILHLQALLGNRWAAIASYLPERTDNDIKNYWNTHLKKKLKKMN
DSCDSTINNGLDNKDPSISNKNTTSHQSSNSSKGQWERRLQTDINMAKQALCDALSIDKP
QNPTNFSIPDLGYGPSSSSSSSTTTTTTTTTTRNTNPYPSPGVYASSAENIARLLQNFMKDT
PKTSVPLPVAATEMAITTAASSPSTTEGDGEGIDHSLFSFNSIDEAEKPKLIDHDINGL
ITQGSLSLFEKWLFDQSHDMIINMSLEGQEVLF*

>G599 (152..1579)
TCGACAGAACAGCTTCGTTGTCACTTGTCTATCTATAAATCGCATCCCCATTGACAACCT
TTCACTTCCACTCAAACCTCTCTCTATATCTCTCTCTCTATATCTCTCTCTATATCT
CTCTCTCTCTTCACTCTCTCTTTCTTTCAAATGGAAAACTCATGGTTCGACATGGAG
ACCCGACCCGGTTTACCGTCCACCGGAAACACCACTCGAACCGATGGAGTTTTAGCTCG
TTCATGGAGCGTCTCTGCTCTCGAAGTCTCCAAGGCTCTAACACCACCAACCCTCAGAT
TCTCTCTCCAAACCGAAGAAGAAGAAGAAGAACCCTCTCTCTGTCTGCTAGACGG
CGACGCGACACGGAAGACACCGGACTTGTCAACCGGAAACCCATTCTCTTCGCTTGTTC
AGAACTTCTCAAATGGTCATGGATCGTATCTTGTCTCACTCTCAAGAAGTATCACCAG
AACATCTGGTTCGGCTATCTCAGTAGTGGTCCACTTAATGGTTCTTTGACCGACAGTCC
TCCTGTGTCTCTCTCCGAATCCGACGACATTAAAGCAATTTTGCAGAGCGAACAATAATC
ATTGAACAGTGTAAATCTCAGTTCGTTCAACGGCGGCAACTCCGGGACCTATAACCGC
TACAGCTACACAGTCCAAGACGGTGGGACGGTGGCTTAAGGACCGGAGAGAGAAAAAGAA
AGAGGAGACTCGGGCTCATAACGCTCAGATTCACGCTGCTGTCTCTGTCGCCGGCGTTGC
TGCAGCTGTTGCTGCTATTGACGAGCCACCGCTGCGTCTTCTAGCTGTGGTAAGGATGA
GCAGATGGCTAAACTGACATGGCCGTTGCTTCTGCTGCGACCCCTGTGGCTGCTCAGTG
TGTGGAAGCTGCTGAAGTTATGGGAGCTGAGAGAGAGTATTGGCTTCTGTTGTTAGCTC
CGCCGTCATGTTCTGTTCTGCGGAGATATTGACTCTCACCGCCGGAGCAGCTACAGC
TTTAAGAGGAGTGCAAAACATTGAAGGCAAGGCAATGAAGGAAGTGTGGAACATAGCATC
AGTGATACCAATGGATAAAGGACTCACTTCTACAGGAGGAAGCAGCAATAATGTTAATGG
TAGCAATGGAAGCTCAAGCAGTAGTCACAGTGGTGAACCTGTACAACAGGAGAATTTCTT
GGGAACCTGTAGTAGAGAATGGCTCGCTAGAGGTTGTGAACCTCCTCAAACGCACTCGCAA
AGGTGATCTCCACTGGAAGATAGTATCTGTTTACATCAACAAAATGAATCAGGTTATGTT
GAAGATGAAGAGCAGGCATGTTGGAGGAACCTTCACCAAGAAGAAAAAGAACATTGTGCT
TGATGTGATCAAGAATGTCCCGGCTGGCCTGGACGACATTTGCTAGAGGGAGGAGATGA
TCTAAGATACTTCGGTTTGAAGACGGTTATGCGAGGTGATGTTGAATTCGAGGTCAAGAG
CCAAAGGGAATATGAATGTGGACACAAGGTGTCTCAAGGCTTCTGTTCTTGTCTGCTGA
GAGGAAGTTTAGGATGTGAATAAACGTTCAATGGCTGCTTGGTTTAAGTGTGAGTTTTTT
TTTAACCTATGTGGTCAAATTTTATTAGTAGGGGTTCTTTTAAGGTAATGGTTTTTTGGG
TTGGGTATAGGATAAAATGGACCTACCAGTCAAGGTGAGGAAGCATTGGGTAAACAAAA
CTTAGTGGGGGTGATCTGTAATATCTATGTTCTTAGTTTTTTTTTGGTTGTTGGTGGTCT
TTTTGTATAAAAAACAAAGTTGAAGTAATAGATATATAGTATGTTTTAATTTTAA

>G599 Amino Acid Sequence (domain in AA coordinates: 187-219, 264-300)
MEKLMVPTWRPDPVYRPPETPLEPMEFLARSWSVSALEVSKALTPPNPQILLSKTEEEEE
EPISSVVDGDGDTEDTGLVTGNPFSFACSETSQMVMDRILSHSQEVSPRTSGRLSHSSG
PLNGSLTDSPPVSPPEDDIKQFCRANKNSLNSVNSQFRSTAATPGPITATATQSKTVGR
WLKDRREKKKEETRAHNAQIHAAVSVAGVAAVAIAAATAASSSCGKDEQMAKTDMAVA
SAATLVAAQCVEAAEVMGAEREYLASVVSSAVNVRSAGDIMTLTAGAATALRGVQTLKAR
AMKEVWNIAVIPMDKGLTSTGGSSNNVNGSNGSSSSSHSGELVQQENFLGTCREWELAR
GCELLKRTRKGLDHLWKIVSVYINKMNQVLMKMSRHVGGTFTKKKKNIVLDVIKNVPAWP
GRHLEGGDDLRYFGLKTVMRGDVEFEVKSQREYEMWTQGVSRLLVLAAERKFRM*

>G804 (114..1139)
ATACTCCAAGAATTTATAGGTTATAAGTAAAAATTCAGTACAAGTTTGTGTTGTTGTTA
TTCCATTTTCTGTGTGTTTTTTTCCCCATAATTTATAAATTTTATAAGCAATATGGAGT
CCCACAACAACAACAGAGCAACAACAACCACTGGTTTCGGCCCATCTGGTCCCATCCA
TGGGACCAATCTCCGGTTCAGTCTCATTAACCACCACTGCTCCAACTCCACTACCACCA
CCGTCACCGCGCTAAACACCCGAAAACGACCGTCCAAGGACCGTCACATCAAAGTAG

ACGGACGTGGCCGAGGATACGTATGCCGGCTATCTGCGCAGCACGTGTCTTCCAATAA
CACGTGAGTTACAACACAAATCGGACGGCGAGACTATAGAGTGGCTGCTCCAACAAGCGG
AGCCAGCTATCATCGCAGCCACCGGAAGTGAACCATACCGCGAATATCTCTACTTTGA
ACATCTCTCTTCGAAGCAGTGGCTCTACTCTTTTTCAGCTCCACTGTCTAAATCTTTCCACA
TGGAAGAGCGGCTCAAACGCTGCCGTTTTTGGGTTCCAGCAACAGCTTTATCATCCTC
ATCATATCACGACAGATTCTTCTTCTTCTTCTTCTTCCCAAAACATTCCGTGAAGAAGATC
TTTTTAAAGATCCTAATTTTCTAGATCAAGAACCCGGTTCAAGATCACCTAAACCGGGAT
CCGAAGCTCCTGATCAAGATCCGGGTTTCGACCCGGTCAAGAACACAAATATGATACCGC
CGATGTGGGCACCTAGCGCCAACGCCAGCCTCCACAAACGGAGGTAGTGCTTTTGGATGT
TACCAGTCGGAGGAGGAGGAGGTCCGGCTAACGTTTCAGGATCCATCACAGCACATGTGGG
CGTTTAATCCGGGTCATTACCCGGGTCGAATCCGGTTCGGTTCAGCTAGGGTCTATGTTAG
TGGGAGGTCAACAGTTAGGGTTAGGTGTTGCAGAAAATAACAATTTGGGGCTATTTCCG
GCGGAGGAGGAGACGGTGGTTCGGGTTGGTCTCGGAATGAGTCTTGAGCAAAAGCCTCAAC
ATCAAGTGAGTGATCATGCTACTAGAGACCAAAATCCTACTATAGATGGTTCCTCTTGAA
AGACTTCATGATTTCTTTGGTTTTTAAAGAGTGTGAATGTGTGATTTATTGCAACTTTTG
TTGAGGACTCCAATGTTAATATGGGTTTTAGGGTTGGCTTTTCGGGATTGCCAAATGTGTT
ATT

>G804 Amino Acid Sequence (domain in AA coordinates: 54-117)
MESHNNQSNNTTGSALVPSMGPISGSVSLTTTAPNSTTTTVAAKTPAKRPSKDRHI
KVDGRGRRIRMPAICARVQLTRELQHKSDGETIEWLLQQAEPALIAATGTGTIPANIS
TLNISLRSSGSTLSAPLSKSFHMGRAAQNAAVFGFQQQLYHPHITDSSSSSLPKTFRE
EDLFKDPNFLDQEPGSRSPKPGSEAPDQDPGSTRSRTQNMIPPMWALAPTASTNGGSAF
WMLPVGGGGGPANVQDPSQHMWAFNPGHYPGRIGSVQLGSMVLGGQQLGLGVAENNNLGL
FSGGGGDGGRVGLGMSLEQKPQHQSVDHATRDQNPTIDGSP*

>G1062 (297..1781)

CAAAAAAAAAAGTTTTCAATTTTTGAAAGCTCTGAGAAATGAAATCTATCATTCTCTCTCTC
TATCTCTATCTTCTTTTTCAGATTTTCGCTTCTTCAATTCATGAAATCCTCGTGATTCTAC
TTTAATGCTTCTCTTTTTTACTTTTCCAAGTCTCTGAATATTCAAAGTATATATCTTTT
GTTTTCAAACCTTTTGCAGAATTGTCTTCAAGCTTCCAAATTTTCAGTTAAAGGTCTCAACT
TTGCAGAATTTTCTCTCTAAAGGTTTCAGACTTTTGGGGTAAAGGTGTCAACTTTGGCGATGG
GTCTTGACGGAACAATGGTGGAGGGGTTTGGTTAAACGGTGGTGGTGGAGAAAGGGAAG
AGAACGAGGAAGGTTTCATGGGGAAGGAATCAAGAAGATGGTTCCTCTCAGTTTAAGCCTA
TGCTTGAAAGGTGATTGGTTTAGTAGTAACCAACCACATCCACAAGATCTTTCAGATGTTAC
AGAATCAGCCAGATTTTCAGATACTTTGGTGGTTTTTCCTTTTAAACCCTAATGATAATCTTC
TTCTTCAACACTCTATTGATTCTTCTTCTTCTTCTTCTTCTTCTCCTTCTCAAGCTTTTAGTCTTG
ACCTTCTCAGCAAAATCAGTTCTTGTCAACTAACAAACAAGGGTTGTCTTCTCAATG
TTCTTCTTCTGCAAAACCTTTTGATAATGCTTTTGAAGTTTGGCTCTGAATCTGGTTTTTC
TTAACCAATCCATGCTCCTATTTTCGATGGGGTTTGGTTCTTTGACACAATTGGGGAACA
GGGATTTGAGTTCTGTTCTGATTTCTTGTCTGCTCGGTCACTTCTTGCCTCGGAAAGCA
ACAACAACAACAATGTTGTGTGGTGGTTTTCACAGCTCCGTTGGAGTTGGAAGGTTTTG
GTAGTCTCTGTAATGGTGGTTTTGTTGGGAACAGAGCGAAAGTTCTGAAGCCTTTAGAGG
TGTTAGCATCGTCTGGTGCACAGCCTACTCTGTTCCAGAAACGTGCAGCTATGCGTCAGA
GCTCTGGAAGCAAAATGGGAAATTCGGAGAGTTCCGGAATGAGGAGGTTTAGTGATGATG
GAGATATGGATGAGACTGGGATTGAGGTTTCTGGGTTGAACTATGAGTCTGATGAGATAA
ATGAGAGCGGTAAAGCGGCTGAGAGTGTTTCAGATTGGAGGAGGAGGAAGGGTAAGAAGA
AAGGTATGCCCTGCTAAGAATCTGATGGCTGAGAGGAGAAGGAGGAAGAAGCTTAATGATA
GGCTTTATATGCTTAGATCAGTTGTCCCCAAGATCAGCAAAATGGATAGAGCATCAATAC
TTGGAGATGCAATTGATTATCTGAAGGAACCTTACAAAGGATCAATGATCTTCACAATG
AACTTGAGTCAACTCCTCCTGGATCTTTGCCTCCAATTCATCAAGCTTCCATCCGTTGA
CACCTACACCGCAAAETCTTTCTTGTCTGTCAAGGAAGAGTTGTGTCCCTCTTCTTTAC
CAAGTCCTAAAGGCCAGCAAGCTAGAGTTGAGGTTAGATTAAAGGAAGGAAGAGCAGTGA
ACATTCATATGTTCTGTGGTCGTAGACCGGGTCTGTTGCTCGCTACCATGAAAGCTTTGG
ATAATCTTGGATTGGATGTTTCAGCAAGCTGTGATCAGCTGTTTTAATGGGTTTGCCTTGG
ATGTTTTCCGCGCTGATGCAATGCAAGAAGGACAAGAGATACTGCCTGATCAAAATCAAAG
CAGTGCTTTTTCGATACAGCAGGGTATGCTGGTATGATCTGATCTGATCCTGACTTCGAGT
CCATTAAGCATCTGTTGAAGCAGAGCTAGAAGAATAAGTCCCTTTAAATCTGCAATTTT
CTTCTCAACTTTTTTTCTTATGTCATAACTTCAATCTAAGCATGTAATGCAATTGCAAAAT

GAGAGTTGTTTTTAAATTAAGCTTTTGAGAACTTGAGGTTGTTGTTGTTGGATACATAAC
TTCAACCTTTTATTAGCAATGTTAACTTCCATTTATGTTTCATCTT
>G1062 Amino Acid Sequence (domain in AA coordinates: 308-359)
MGLDGNNGGVWLNNGGGEREENEESWGRNQEDGSSQFKPMLLEGDFSSNQPHQDLQM
LQNQPDFRYFGGFPPFNPDNLLQLHSIDSSSSCSPSQAFSLDPSQQNQFLSTNNKGCCLL
NVPSSANPFDNAFEFGSESGFLNQIHAPISMFGFSLTQLGNRDLSSVPDFLSARSLLAPE
SNNNTMLCGGFTAPLELEGFGSPANGGFVGNRAKVLKPLEVLASSGAQPTLFQKRAAMR
QSSGSKMGNSESSGMRRFSDDGMDDETGIEVSGLNYESDEINESGKAAESVQIGGGKGK
KKGMPAKNLMAERRRRKKLNDRLYMLRSVVPKISKMDRASILGDAIDYLKELLQRINDLH
NELESTPPGSLPPTSSSFHPLTPTPQTLSCRVKELCPSSLPSKQQAARVEVRLREGRA
VNIHMFCCRPPGLLLATMKALDNLGLDVQQAIVISCFNGFALDVFRAEQCQEQEILPDQI
KAVLFDTAGYAGMI*

>G1322 (213..833)
AAAGTTATTGATAGTTTCTGTTACTTATTAATTTTAAAGGTTATGTGTATTATTACCAAT
TGGAGGACTATATAGTCGCAAGTCTCAACCCTATAAAAGAAAACATTCGTCGATCATCTT
CCCGCCTCGAGTATCTCTCTCTCTCTCTCTCTCTGTTTTCTTTATTGATTGCATAGA
CAAAAATACACACATACACAACAGAAAGAAAGATGGAGACGACGATGAAGAAGAAAGGGA
GAGTGAAAGCGACAATAACGTCACAGAAAGAAGAAGGAACAGTGAGAAAAGGACCTT
GGACTATGGAAGAAGATTTTCATCCTCTTTAATTACATCCTTAATCATGGTGAAGGTCTTT
GGAACCTCTGTGCGCAAAGCCTCTGGTCTAAAACGTACTGGAAAAAGTTGTGCGCTCCGGT
GGCTGAACATATCTCCGACCAGATGTGCGGCGAGGGAACATAACCGAAGAAGAACAGCTTT
TGATCATTCAGCTTCATGCTAAGCTTGGAACAGGTGGTGAAGATTGCGAAGCATCTTC
CGGGAAGAACGGACAACGAGATAAAGAACTTCTGGAGGACAAAGATTGAGAGACACATGA
AAGTGTGATCGGAAAATATGATGAATCATCAACATCATTGTTGCGGAAACTCACAGAGCT
CGGGGATGACGACGCAAGGCAGCTCCGGCAAAGCCATAGACACGGCTGAGAGCTTCTCTC
AGGCGAAGACGACGACGTTTAAATGTGGTGGAAACAACAGTCAAACGAGAATTACTGGAACG
TTGAAGATCTGTGGCCCGTCCACTTGCTTAATGGTGACCACCATGTGATTAAAGATATAT
ATATAGACCTCCTATACATTTATATGCCCCAGCTGGGTTTTTTTTGTATGGTACGTTATTT
GGTTTTTCTATTGCTGAAATGTCGTTGCTTTAATTTACATACGAAAAGTGCATTAAATC
ATTAAATCTTCAATACATATGGAGGTGGTGTGTTGAGTAAAAAAAAAAAAAA

>G1322 Amino Acid Sequence (domain in AA coordinates:26-130)
METTMKKKGRVKATITTSQKEEGTVRKGPWTMEEDFILFNYILNHGELWNSVAKASGLK
RTGKSCRLRWLNLYLRPDVRRGNITBEEQLLIQLHAKLGNRWSKIAKHLPGRTDNEIKNF
WRTKIQRHMKVSSSENMNHQHCSGNSQSSGMMTQGSSEKAIIDTAESFSQAKTTTFNVVE
QQSNENYWNVEDLWVPVHLLNGDHHVI*

>G1331 (1..786)
ATGGTGGAGAAGTTTGGAGAAAGGGTCCATGGACCCGCGAAGAAGACCGTCTTTTGATC
GAATACGTCCGTGTTACGGTGAAGGTCGTTGGAACCTGTCTCTAAACTCGCAGGATTG
AAAAGGAATGGCAAAGCTGCAGACTAAGATGGGTGAATTACCTTAGACCAGACCTCAAG
AGAGGACAGATCACTCCACATGAAGAAAGTATAATACTTGAGCTACACGCTAAGTGGGGA
AATAGGTGGTCAACAATTGCACGTAGTTTACCAGGAAGAACAGACAATGAGATCAAGAAC
TATTGGAGAACCATTTCAGAAAAAGGCAAAGCCTACGACTAACAATGCGGAGAAGATA
AAGAGTCGTCTCCTAAAAAGGCAACACTTCAAGGAACAGAGAGAAATAGAGTTGCAACAA
GAACAGCAGTTGTTTCAGTTTCGACCAACTCGGTATGAAAAAGATCATCTCTTTACTCGAA
GAAAACAATAGCAGTAGCAGTAGCGATGGCGGTGGTGTGTTCTATTATCCTGATCAA
ATAACACATTTCATCAAAACCCTTTGGCTATAACTCTAATTCATTAGAGGAGCAGTTACAA
GGTAGATTTTCTCCTGTAAACATACCTGATGCTAATACTATGAACGAAGACAATGCCATA
TGGGACGGGTTTGGAAACATGGATGTTGTAAATGGACATGGTGGGAACTGGGTGTTGTG
GCTGCTACTGCTGCTTGTGGCCCAAGGAAGCCCTATTTCCATAACTTGGTGATTCCATTT
TGTTAA

>G1331 Amino Acid Sequence (conserved domain in AA coordinates:8-109)
MVVEVWRKGPWTAEDRLLIEYVRVHGEGRWNSVSKLAGLKRNGKSCRLRWVNYLRPDLK
RGQITPHEESIILELHAKWGNRWSTIARSLPGRTDNEIKNYWRTHFKKKAKPTTNNAEKI
KSRLKQRHFKEQREIELQEQQLFQFDQLGMKKIISLLENNSSSSSDGGGDVFFYPDQ
ITHSSKFPGYNSNSLEEQLQGRFSPVNIPDANTMNEEDNAIWDGFWMMDVNVNGHGNLGVV
AATAACGPRKPYFHNLVIPFC*
>G1521 (1..891)

ATGCCTCCATTACCGTCTCCACGGCGCCTTCGTCTTCGAGACATCTTCGATCGCCGAA
AGTATCGCGAAATTTGCAGGGAGAGCAATATTTCTGCTTTACAGGGGAAATCGTGCCG
ATATGCCTCGAAAATCTAACCGAGCGAAGATCCGCCGCCGTGATCAGGTTGTGCAAGCAC
GGATACTGCCTTGCTTGTATTTCGGAAGTGGAGCAGCTTCAAGAGGAATTGTCCTCTTGT
AACACTCGTTTTGATTCTCGTTTATCGTTAGTGATTTTGCTTCTAGAAAATACCATAAG
GAGCAATTACCAATTCTTCGTGATCGTGAGACTTTAACTTATCATCGGAATAATCCTTCC
GATCGCCGGAGGATAATTCAAAGGTCGAGGGATGTTTTGGAAAACCTAGCTCAAGATCA
AGGCCATTGCCATGGCGGAGATCATTTGGACGACCAGGTTCACTTCCTGATTCTGTTATC
TTCCAGCGAAAGCTTCAGTGGCGAGCTAGCATATACACTAAGCAATTACGAGCTGTTCTGA
TTACATTCAAGGCGCTTGGAACTAAGTTTGGCGGTGAATGATTACACCAAAGCAAAGATA
ACTGAAAGAATTGAGCCATGGATTAGAAGAGAGCTTCAGGCAGTCTTGGAGATCCTGAT
CCCTCAGTTATTGTTTCACTTTGCGTCAGCTCTTTTCATCAAAGGCTTGAGAGAGAGAAT
AATCGACAAACCGGGCAGACCGGGATGTTGGTGAAGATGAAGTCTCCTCTCTTCGAAAA
TTCTTGTCTGATAAAGGTGGATATATTTTGGCATGAACATAAGATGTTTTGCGGAGAGTATA
CTCAGGATGGAGACTTATGATGCAGTGGTTGAATACAATGAGGTGGAGTAA

>G1521 Amino Acid Sequence (domain in AA coordinates: 39-80)
MPPLPSSSTAPSSSRHLRSPESIAKFAGRAIFPALQKSCPICLNLTERRSAAVITVCKH
GYCLACIRKWSSFKNRNCPLNTRFDSWFIVSDFASRKYHKEQLPILRDRELTLYHRNPS
DRRRRIQSRSDVLENSSSSRPLPWRRSFGRPGSVDPDSVIFQRKLQWRASIYTKQLRAVR
LHSRRLLESLAVNDYTKAKITERIEPWIRRELQAVLGDPPSVIVHFASALFIKRLEREN
NRQTGQTGMLVEDEVSSLRKFLSDKVDIFWHELRCFABSILTMETYDAVVEYNEVE*

>G183 (1..1458)

ATGAGTGATTTTGATGAAAACCTTCATCGAAATGACGTCGTATTGGGCTCCACCATCCAGT
CCTAGCCCAGAACGATATTGGCAATGCTGGAGCAAACCGACAATGGTCTGAATCCAATC
AGTGAGATCTTCCCTCAAGAAAGCTTGCCAAGAGATCATACTGATCAATCTGGACAAAGA
TCTGGTCTTCGTGAGAGACTGGCTGCAAGAGTAGGATTCAATCTTCCAACACTCAATACA
GAAGAAAACATGAGTCTTTGGATGCATTTTTCAGGAGCTCGAATGTTCTTAATTCCTCCT
GTCGTTGCAATCTCTCCAGGATTCAGTCCATCAGCACTATTGCATACTCCCAATATGGTC
AGTGATTCTTCCAGATTATCCCTCCGTCTTCAGCCACCAATTACGGACCTCTAGAGATG
GTGGAACCTTCCGGTGAAGACAATGCACGATGATGATGTTCAACAACGATCTTCTTAT
CAGCCGTACAATGTTGATCTGCCCTTCTCTAGAAGTCTTTGATGATATTGCAACGGAAGAG
TCCTTTTATATCCCATCTTATGAACCTCATGTTGACCCAATTGGAACCTTTAGTCACA
TCCTTTGAATCTGAACTCGTTGACGATGCCCATACCGACATCATCTCCATTGAGGACAGT
GAGAGCGAGGATGGAAACAAAGATGATGACGACGAGGACTTCCAATACGAAGACGAAGAC
GAAGACCAATACGACCAAGATCAAGATGTAGATGAAGATGAAGAGGAAGAAAAAGATGAA
GACAATGTTGCATTAGATGATCTCAACCTCCACCTCCAAAGAGAAGGAGATATGAGGTA
TCAAACATGATTGGAGCCACAAGAACAAGACAAGACAAGGATCATACTTCAGATGGAA
AGCGACGAAGACAATCCTAACGATGGTTATCGTGGAGAAAATACGGTCAGAAAGTCGTC
AAAGGAAATCCTAATCCGAGGAGTTACTTCAAGTGACAAACATCGAGTGCAGAGTGAAA
AAACATGTGGAGAGAGGAGCAGACAATATCAAGTTGGTTGTGACTACATACGATGGGATA
CACAACCATCCTTCACCACCTGCACGTAGAAGCAATTCAGTTCAAGGAACCGGTCTGCA
GGGGCAACAATACCTCAAATCAGAATGATCGAACCAGTCGGTTAGGTAGGGCTCCTCCT
ACTCCTACTCCTCCTACTCCTCCTCCTCGTCTTACACACCTGAGGAGATGAGGCCTTTC
TCTTCGTTGGCTACAGAAATTGATCTGACAGAGGTTTATATGACCGGAATCTCTATGCTG
CCGAATATACCGGTTTACGAGAATTCGGGTTTTATGTACCAGAATGATGAACCGACGATG
AATGCGATGCCGATGGTTAGATGTGTACGATGGGATCATGGAACGCCTGTATTTTAAG
TTTGGTGTGACATGTAG

>G183 Amino Acid Sequence (domain in AA coordinates: TBD)
MSDFDENFIEMTSYWAPPSSPSRPTILAMLEQTDNGLNPISBIFPQESLPRDHTDQSGQR
SGLRERLAARVGNLPTLNTEENMSPLDAFFRSSNVPNSPVVAISPFGFSPSALLHTPNMV
SDSSQIIPPSSATNYGPLEMVETSGEDNAAMMMFNNDLPYQPYNVLDLPSLEVFDDIATEE
SFYIPSYEPHVDPIGTPLVTSFESELVDDAHTDIISIEDSESEDGNKDDDDDFQYEDD
EDQYDQDQDVEDEEEKDEDNVALDDPQPPPPKRRRYEVSNMIGATRTSKTQRIILOME
SDEBNPDGYYRWRKYGQKVVGKPNPNRSYFKCTNIECRVKHVERGADNIKLVVTTYDGI
HNHPSPPARRSNSSSRNRSAGATIPQNQNDRTSRLGRAPPTPTPTPPSSYTPEMRPF
SSLATEIDLTEVYMTGISMLPNIPVYENSGFMYQNDEPTMNAMPDGSVDYDGMERLYFK
FGVDM*

>G2555 (177..956)

CTGTTTTTGTATCCGTGTAAATTAATCACACGGTAGTTTTTGTATGAAAAGACAACAATCG
GAGAACAACTCTGGTCTGCTGCTAAAATTTAATAAATTTGTTTTGTCTAATTGTCTCCACCC
ATAAAAAAGCGCGAATTCAATTCACCGACTAAAGACATTCTCCGGTGGAGACCCCGATGC
AATCCACTCATATAAGCGCGGAAGTAGCGGTGGTGGTGGAGGAGGAGAGGTGA
GTCGAAAGTGGATTATCTCGGATCCGTTCCAGCTCCAGCTACTTGGATTGAAACCCTACTCG
AAGAAGATGAAGAAGAAGGTTTAAAACCTAACCTTTGTTTAAACAGAGCTGCTTACTGGTA
ATAATAACTCTGGAGGAGTGATAACGAGTCGTGACGACTCGTTCGAGTTCCTGAGTTCCTG
TTGAGCAAGGATTGTATAATCATCATCAAGGTGGTGGCTTTCACCGTCAGAATAGTTCCTC
CGGCTGATTTTCTTAGTGGGTCTGGTCTGGGACTGATGGGTATTTCTCTAATTTTGGTA
TTCCGGCGAATTATGACTATTTGTGCGACCAACGTTGATATTTCTCCGACTAAACGGTCTA
GAGATATGGAAACACAGTTTCTTCTCAGCTGAAAGAAGAGCAAATGAGTGGTGGGATAT
CAGGAATGATGGATATGAACATGGACAAGATTTTGTAGGATTCAAGTTCCTTGAGGGTTC
GTGCTAAACGTGGTTGTGCTACTCATCCTCGTAGCATTTGCTGAACGGGTGAGAAGAACGC
GAATAAGTGATCGGATTAGGAGGCTGCAAGAGCTTGTTCCTAACATGGATAAGCAAACCA
ACACTGCAGACATGTTGGAAGAAGCTGTGGAGTATGTGAAGGCTCTTCAAAGCCAGATCC
AGGAATTGACAGAGCAGCAGAAGAGATGCAAATGCAAACCTAAAGAAGAACAATAATGTA
TCCTTTAGGATTTGATATATCTGTATTTTATTTTGTACTATCTAAAAATGGTGATGATC
TGTTTCGAAAATTCGAAACATGATCTTATATATTGAAGTAAAGAAATAGATATATATGAA
TTTTAGCTGTAAATTTTGTACAATAAGGAGAAAAAGATTTAGAAGAGTCAATAAAAAAG
ATGATGTTTACAAGTCAAAAAA

>G2555 Amino Acid Sequence (domain in AA coordinates: 175-245)

MQSTHISGGSSGGGGGGGGEVSRSLRSRISAPATWIEILLLEDEEEGLKPNLCLTELLT
GNNNSGGVITSRDSFEFLSSVEQGLYNHHQGGGFHRQNSSPADFLSGSGSGTDGYFSNF
GIPANYDYLSTNVDISPTKRSRDMETQFSSQLKEEQMSGGISGMMDMNMDKIFEDSVPCR
VRAKRGKATHPRISIAERVRRTRISDRIRRLQELVPNMKQNTADMLEEAVEYVKALQSQ
IQELTEQQKRCKCKPKEEQ*

>G375 (53..1171)

TCGACAAAACTCTCACTCTCCCTCAAACATAACATAACAGAACACAAAATGGGTCT
CACTTCTCTTCAAGTTTGCATTTCTGATTGGCTCCAGGAATCCGAGTCATCAGGAGG
AAGCATGTTAGACTCTTCAACGAATTCTCCGTCAGCAGCCGACATACTAGCAGCTTGCAG
CACTAGACCACAAGCCTCGGCCGTGGCTGTAGCCGCTGCAGCTCTGATGGACGGTGGAAAG
GAGGCTGCGTCCACCTCACGACCATCCTCAAAGTGTCTCGTTGCGAGTCAACACATAC
TAAGTTCTGTTACTACAATAACTACAGCCTCTCTCAGCCTCGTTACTTCTGCAAGACTTG
TCGCCGTTACTGGACAAAAGCGGAACTCTAAGGAATATTCCGGTGGTGGTGGATGCCG
TAAAAACAAGAAACCATCTTCTCTAATTCCTCCTCCTCCACTTCTTCCGGCAAAAAACC
ATCCAACATCGTTACCGCCAATACCTCTGATCTTATGGCTTTAGCACATTCTCATCAAAA
TTACAACATCTTCTCTCAGGTTTTCACATTTTGGTGGGATGATGGGGTCTTACTCAAC
TCCGGAGCATGGTAACGTTGGTTTCTTGGAGAGCAAGTATGGCGGTTTGCTTTCGCAGAG
CCCTAGACCTATTGATTTCTTGGACAGTAAGTTTGATCTCATGGGAGTGAACAATGACAA
CCTGGTCATGGTTAATCATGGAAGTAACGGAGATCATCATCATCATAATCATCATCAT
GGGTCTGAATCACGGTGTAGGTCTTAACAACAACAACAACATGGTGGATTTAATGGGAT
TTCTACGGGAGGCAATGGAATGGTGGTGGTCTCATGGATATATCGACATGCCAAAGACT
TATGCTATCTAATTATGATCATCACCATTACAATCATCAAGAAGATCATCAAAGGGTAGC
AACAATAATGGATGTGAAGCCAAATCCGAAGTTGTTATCGCTTGATTGGCAGCAAGATCA
ATGCTACTCCAATGGTGGTGGTAGCGAGGCGAGGAAAATCCGACGGTGGTGGATACGG
CAATGGTGGTTATATCAACGGTTTAGGTTCTGTCGTGGAATGGTTTGATGAATGGCTATGG
AACGTCCACTAAAAAACTCCTTGGTTTGATAAGTTAATCAGAACTTCTTTTCTTGT
CGTCATCAACTAGTAGTAGTAATAGTAGTTGGAGACTAGAGAAGCACTTCAAATTAT
TTATGGGTTTGTGTTGCTAAGCCAGTTTAC

>G375 Amino Acid Sequence (domain in AA coordinates: 75-103)

MGLTSLQVCMDSWLQESSESSGSMDSSTNSPSAADILACSTRPQASAVAVAAAALMD
GGRRLRPHPDHQKPCRESTHTKFCYNNYSLSQPRYFKTCRRYWTGKGTLRNIPVGG
GCRKNKPPSSSSSSSGKPSNIVTANTSDLMALAHSHQNYQHSPLGFSHFSGMMGS
YSTPEHGNVGFLESKYGGLLSQSPRPIDFLDSKFDLMGVNNDNLVMVNHGSNGDHHHHH
HHMGLNHGVGLNNNNNGGFNGISTGGNGNGGGLMDISTCQRLMLSNDHHHYNHQBHQ
RVATIMDVKNPKLLSLDWQDQCYSNGGSGGAGKSDGGGYNGGYINGLSSWNGLMN

>G1007 (86..763)

ATTCCTTCTTGCCCTAGGAACATAATTGTTGCACACTTCGGTACACAATTTTTTGAGCACTT
 CGCATCAAAACGAGAGAGAGAAAAGAATGGTGGATTCTCATGGCTCCGACACGGAATGTTT
 CTCCAAGAAGAAAAAGGAGAAAAACGAAAGAAAAGGGGGTATATCGTGGGGCTCGCATGAG
 GAGCTGGGGGAAATGGGTCTCGGAGATTGGGAGCCCCGTAAGAAATCAAGAATCTGGCT
 CGGGACTTTCCCCACGCGGGAGATGGCAGCGCGTGCCCATGATGTTGCGGCATTGAGTAT
 CAAAGGAAGTTCGCAATCCTTAACTTCCTTGAGCTCGCGGATTTTCTGCCAAGACCGAT
 CTCGCTCAGCCAAACAGGATATTCAGGCCGAGCGCGCGAAGCCGCTCTTATGGATTTTCAA
 AACTGTACCAATTCATCTTCAGGATGACTCAACGCGGTGTCAAACTAGGTGTGATACTGA
 AAGATGCGAAAAGTGGTCTATCCTCATCGTCTCAGCCTCATCCTCATCCTCATCTTCGTC
 CTCGTCCTCATCATCTATGCTTTTCGGGGGAGCTAGGAGATATTGTGGAGTTGCCGAGTCT
 TGAAAACAATGTAAATACGATTGTGCGCTGTATGACTCGTTGGAGGGGGCTGGTGTGAT
 GCCCCATGGTTAGATGCTACCGAAAATGATTTTAGGTATGGAGATGATTCCGGTACTGTT
 GGACCCATGTCTCAAAGAAAGCTTTTTGTGGAATTATGAGTAAGGTTTTTTTTTGGAAAG
 AAATGTGGTTTTTTTGTTCCTCCTCTCTTTTATACTTTCGATCTTTTTTCTTAAGCATAT
 ATATCTTACATATGTAATACTTTTCCATTAGTAACAATGATTCGGTTTCGGGTACAA
 AAAAAAAAAAAAAAAAAAAAAA

>G1007 Amino Acid Sequence (domain in AA coordinates: TBD)

>SG1007 AMINO ACID SEQUENCE (COMPLIN 1X IN FOLDING)
MVDSHGSDTECCSSKKKKKEKKEKGVYRGMRNRSWGKWNVEIREPRKKSRIWLGTFPTAEM
AARAHDAVALSICKSSAILNFPPELADFLPRPVSLSQODIQAAAAEAALMDFKTVPFHLQD
DSTPLQTRCDTEKIEKWSSSSSSSSASSSSSSSSSSSSSSMLSGELGDIVELPSLENNVKYDC
ALYDSLEGLVSMPPWLDATENDFRYGDDSVLLDPCLKESFLWNYE*

>G1010 (344..1276)

ATTCTTCTTGTAGAAAACCTTGACAACCTTTTGTGTTTTGTTTTCTTTCTTCTGTAATTTTTT
AAAAGAGAGAGAGCTATGTAGCTATGAAACAGTAAGAGATATAGATATAGAGAGACAGAG
AAAGATGATGATCAGTGAAGTTAGGCTAAACCCACTTTCTATTTATGTATAATTAGGTCA
ATCACATCACCAATCTCCTCCTCCAATCTCCTCCTCTCCTTCCAAATTCTAGGGTTTTG
CTTGATATCTACCCCCCTTTCTCAATTCCCTAGGGAAACTGTGAATTTTCATCAAATTCAT
TATTTTTTTGGTCCACACCCTTAAAGAGATCTGAGAGTCTTAAAGATGATGACAGATTTATC
TCTCAGCAGAGATGAAGATGAAGAAGCAAGCAAGCCCTAGCAGAGAAGAAGGAGCGCG
TGAAGTAGCAGACAGAGAGCAATGTTGCGACAAGGTTGTGACTCCAAGTGATGTCGGAAA
ATAAACCAGCATTTGTGATCCCAAAGCAACACGCAGAGAGATTCTTCCCTTTAGATTTCATC
TTCAAACGAGAAAGGTTTTGCTTTTAAACTTCGAAGATCTCACTGGCAAATCTTGGAGGTT
CCGTTACTCTTACTGGAACAGTAGTCAAAGCTATGTGATGACTTAAAGGTTGGAGCAGATT
CGTTAAAGACAAAAAGCTTGACGCCGGAGATATTGTCTCTTTCCAAAGATGTGTCCGGAGA
TTCAGGAAGAGATAGCCGTTTTGTTTATTGATTGGAGGAGAAAGACCTAAAGTCCCTGACCA
TCCTCATTTTCGCCGCCGGAGCTATGTTCCCTAGGTTTTACAGCTTTCCCTCGACCAATTA
CAGTCTTTTATAATCATCAGCAGCAACGTCATCATCAGTGGTGGTGGTTATAATTATCA
TCAAATTCGAGAGAAATTTGGTTATGGTTACTTCGTTAGGTCAGTGGATCAGAGGAACAA
TCCTGCGGCTGCGGTGGCTGATCCGTTGGTGATTGAATCTGTGCCGGTGATGATGCACGG
GAGAGCTAATCAGGAACCTTGTTGGAACGCCCGGGAAGAGACTGAGGCTTTTTTGGAGTTGA
TATGGAATGCGGCGAGAGCGGAATGACCAACAGTACGGAGGAGGAATCATCATCTTCCGG
TGGAAGTTTGCCACGTGGAGGCGGTGGTGGTGCTTCATCTTCTCTTTCTTTCAGCTGAG
ACTTGGAAGCAGCAGTGAAGATGATCATCTTCACTAAGAAGGAAGTCTTTCATTGTCTTT
TGATTGTGATCAATAATAATGATGATGAAATAGTTGGTATTTTAAAGAAAAAAACA
TACATTATATAATCTTATATATATGACAACATAATGCATTGATTTCCCTT

>G1010 Amino Acid Sequence (domain in AA coordinates: 33-122)

NNTDLSLRTRDEDEEEAKPLAAEEGAREVADREHMFDFKVVTPSDVGKLNRLVLPKQHAERF
 FPLDSSSNEKGLLLNFEDLTGKSWRFYRSYWNSSQSVMYTKGWSRFVKDKKLDA GDIVSF
 QRCVGDSDGRDSRLFIDWRRRPKVPDHPHFAAGAMFPRFYSFPSTNYSLYNHQQQRHHHSG
 GGYNYHQIPREFGYGYFVRSVDQRNNPAAAVADPLVIESVPVMMHGRANQELVGTAGKRL
 RLFGVDMCEGSEGMTNSTEEESSSSGGSLPRGGGGGASSSSFFQLRLGSSSEDDHFTTKG
 KSSLSDFLDL*

>G1014 (174..1112)

CAG1014 (1747..11112)
CAG1014AACACAGTCTCTTTCTCTCTATCTATCTTCTCTTTCTCTCTATCTCTAT
CACTGAAACCCAAAGATCCACCATTGTGTTCTTTTTCCTTCACACAGAGAACTGTTTT

CTTCCACACTTCCTTTTACTAGGCAGTGTTAAACCAATTGAGAGAGAAAAATGATGGTTG
ATGAAAATGTGGAACCAAGGCCTCTACTTTAGTGGCAAGTGTGATCATGGGTTTGGAT
CCGGGTCCGGTTCATGATCATCATGGGTTATCGGCGTCTGTGCCTCTTCTTGGTGTAACT
GGAAGAAGAGAAGGATGCCTAGACAGAGACGATCTTCTTCTTCTTTAACCTTCTCTCTT
TCCCTCCTCCTATGCCTCCTATTTTCCCACGTGCCAACTCCTCTCCCCGCACGTAAAAATTG
ACCCAAGAAAGCTAAGATTCTCTTCCAAAAGGAACTCAAGAACAGTGACGTCAGCTCTC
TCCGACGTATGATACTCCCGAAGAAAGCCGCGGAGGCTCACTTGCCGGCACTTGAATGCA
AGGAAGGGATTCCCTATAAGAATGGAAGATTTGGACGGTTTTTCACGTTTGGACCTTCAAGT
ATAGGTACTGGCCAAACAACAATAGCAGAATGTACGTGCTAGAAAAACACAGGCGATTTTG
TGAATGCTCATGGTCTGCAGCTAGGTGACTTCATCATGGTTTACCAAGATCTCTACTCAA
ACAATTACGTTATACAAGCAAGAAAAGCATCGGAAGAAGAAGTAGACGTAATCAATC
TTGAAGAAGACGACGTTTACACAACTTAACAAGGATCGAAAACACTGTGGTTAACGATC
TTCCTCCTCCAAGATTTTAATCATCACAAACAACAACAACAACAACAGCAACAGCAACA
GCAACAAATGTTCTTACTATTATCCAGTCATAGATGATGTACCCACAAACACAGAGTCTT
TTGTCACGACACGACGGCTCTTACCTCCAACGATACTCCTCTCGATTTTTTGGGTGGAC
ATACGACGACTACTAATAATTATTACTCCAAGTTCGGAACATTCGATGGTTTGGGCTCCG
TTGAGAATATCTCTCTCGATGACTTCTACTAGATAATCAATCGATGGGCTCATGGTATTC
TTGATGGTGATCAGCTATTTAATATCCTTATAATATATATAAGAATTAAATGCAATTTGC
ATATATATTATCAAGTGTGTGAATATAACATTACAGTTTAAAAAAAAAAAAAAAAAAAA

>G1014 Amino Acid Sequence (domain in AA coordinates: 90-172)
MVDENVETKASTLVASVDHGFSGSGHDHGLSASVPLLGVNWKRRMPRQRRSSSSFNL
LSFPPPMPPISHVPTPLPARKIDPRKLRFLFQKELKNSDVSSLRRMILPKKAAEAHLPAL
ECKEGIPIRMEDLDGFHVWTFKYRYWPNNNSRMYVLENTGDFVNAHGLQLGDFIMVYQDL
YSNNYVIQARKASEEEVDVINLEEDDVYTNLTRIEN TVVNDLLLQDFNHHNNNNNNNSN
SNSNKCSYYPVIDDVTTNTESFVYD TTALT SNTDPLD FLGGHTTTTNNYYSKFGTFDGL
GSVENISLDDFY*

>G1035 (103..624)
CCATAATAATATATTAATACTATATACTATAATCTTTTACATAATAAACTTTGGGTCCT
GCGTCTTAATCATAGTACTTAATTTTCTCTGTGTGTTTAAATATGAATAATAAACTGAA
ATGGGATCTTCCACAAGTGGAAATTGCTCGTCCGTTTCAACCACTGGTTTAGCTAACTCC
GGTTCAGAATCTGATCTCCGGCAACGTGATCTAATCGACGAGCGGAAGAGAAAGAGGAAA
CAGTCGAACAGAGAATCTGCGAGGAGGTCGAGGATGAGGAAGCAGAAGCATTGATGAT
CTCACTGCTCAGGTGACTCATCTACGTAAAGAAAACGCTCAGATCGTCGCCGGAATCGCC
GTCACGACGCAGCAGCTACGTCACTATCGAGGCGGAGAACGACATTCTCAGAGCTCAGGTT
CTTGAACCTTAACCACCGTCTCCAATCTCTTAACGAGATCGTTGATTTCTGTCGAATCTTCT
TCTTCAGGATTCGGTATGGAGACCGGTCAGGGATTATTCGACGGTGGATTTATTCGACGGC
GTGATGAATCCTATGAATCTAGGGTTTTATAATCAACCAATCATGGCTTCTGCTTCTACT
GCTGGTGATGTTTTCAACTGTGTAGAAAACCTCACATCATTATCATCGTGAGTGAGACTAA
TCATCGCAGCAGGGGTAAACTGTAATTTTCTTATAAATTATGTGATGATGCTTTGTTT
CTTTATTTTATAAGATGGTTAATTAGTGTTTAAACTGATTGTAATGATAGACAGTGTAA
GAAATGTGTGATATCATGGAGATGGTGATGTGAGTTTGGTACAAATATTTAAGATCTTT
TCTTTCTATATATTAAGTGAAGAAATAATATTTTGTCTATTTCTTAAAAAAAAAAAAA
AAA

>G1035 Amino Acid Sequence (domain in AA coordinates: 39-91)
MNNKTEMGSSSTSGNCSSVSTTGLANSSES DLRQRDLIDERKRKRKQSNRESARRSRMRK
QKHLDDLTAQVTHLRKENAQIVAGIAVTTQHYVTIEAENDILRAQVLELNHRLQSLNEIV
DFVESSSSSGFMETGQGLFDGGLFDGVMNPMNLGFYNQPI MASASTAGDVFNC*

>G1046 (1..567)
ATGATTAGACATCTAAAACCCCTACATGGAGTCGTCTAGTGTCATCGCTCTCATTGTTTT
GATATTCCTTGATGGAGTCCCACTACACGACGATCATTTCAACTCGGCATTCTACCAAAC
ACTGACTTTAATGTTTCAATTTGCAGTCAAACGTATCGACCCGATCAACAATCAGTCTCAC
TTAGACCCAAATGCAGAAAACATTTCCATAACGAAGGCTTTGCTCCAGAAGAAAGAAGA
GCAAGAAGAATGGTCTCTAACCGGAATCTGCAAGGAGGTCAGTATGCGCAAAAAGAAG
CAGATCGAAGAGCTGCAACAACAAGTTGAACAACATCATGATGTTGAATCATCACTGTCT
GAGAAAGTCATCAACTTGTGGAAAGCAACCATCAGATCCTACAAGAGAACTCACAGCTG
AAAGAGAAAGTCTCTTCTTTCACTTGCTCATGGCAGATGTGCTATTACCATGAGAAAT
GCAGAGAGCAACATCAATGACCGCAATGTGAATTATCTAAGAGGAGAACCATCAAACCGT

CCCACCAACAGTCCCTTTGGTAAGTAA

>G1046 Amino Acid Sequence (conserved domain in AA coordinates:79-138)

MIRHLKPYMESSSVHRSHCFDILDGVPLHDDHFNSAFLPNTDFNVHLQSNVSTRINNQSH
LDPNAENIFHNEGLAPEERRARRMVSNNRESARRSRMRKKQIEELQQQVEQLMMLNHHLS
EKVINLLESNHQILQENSQLEKQVSSFHLLMADVLLPMRNAESNINDRNVNLYLRGEPSNR
PTNSPFGK*

>G1049 (29..550)

CTAACTTTCTTCCCAAGTAAACTTCAAAATGCAGCCGCAACAGACGTTTTCAGCCTCCA
TAACTACCTAAACTCATCGATACTGCAGTCTCCGTATCCTTCTAATTTCCCGATATCTAC
GCCATTTCCAACCAACGGTCAAAACCCGTACCTCCTCTACGGATTCCAAGCCCTACAAA
CAATCCACAATCCATGAGCCTAAGCAGCAACAACATCAACATCAGATGAAGCAGAAGAGCA
GCAGACGAACAACAATATAATCAACGAGCGGAAGCAGAGAAGGATGATTTCAAACCGAGA
ATCCGCAAGGAGATCGCGTATGAGGAAGCAAAGACACCTTGACGAGCTTTGGTCACAAGT
GATGTGGTTAAGGATCGAGAATCATCAGTTGCTTGATAAGCTTAACAATCTCTCTGAGTC
TCACGACAAGGTTCTTCAAGAGAATGCTCAGCTTAAAGAAGAAACATTTGAGCTTAAGCA
AGTGATCAGCGATATGCAAATTCAAAGCCCTTTCTCTTGCTTTAGAGACGATATAATCCC
CATTGAATAAAGCATTTTCCCGATTTCATATTTATGAAAATTTTCTTCAAGAGTATGTT
TCTTTGTATGTATATGTGGAGATGTATTCAGGGTTTTGATAATATGACCCTTTACGACG
ACGTTTTTTAGATTGTAGTAAATTTATAAACTAAAGAAGATTAGTGTTAATGAAGAACAAA
TATAA

>G1049 Amino Acid Sequence (domain in AA coordinates 77-132)

MQPQTDVFLHNYLNSILQSPYPSNFPISPTPPTNGQNPYLLYGFQSPNTNPQSMSSLSS
NNSTDEAEQQTNNNIINERKQRRMISNRESARRSRMRKQRHLDELWSQVMWLRIENHQ
LLDKLNNLSHSDKVLQENALKEETFELKQVISDMQIQSPFSCFRDDIIPB*

>G1069 (89..934)

TTGGAACCCTAGAGGCCCTTCAAGCAAATCATCAGGGTAACAATTTCTTGATCTTTCTTT
TTAGCGAATTTCCAGTTTTTGGTCAATCATGGCAAACCTTGGTGGACGAACCAGAGTGG
TTTAGCGGGCATGGTGGACCATTCGGTCTCCTCAGGCCATCACCAAAACCATCACCACCA
AAGTCTTCTTACCAAAGGAGATCTTGGAAATAGCCATGAATCAGAGCCAAGACAACGACCA
AGACGAAGAAGATGATCCTAGAGAAGGAGCCGTTGAGGTGGTCAACCGTAGACCAAGAGG
TAGACCACCAGGATCCAAAAACAACCCAAAGCTCCAATCTTTGTGACAAGAGACAGCCC
CAACGCACTCCGTAGCCATGTCTTGGAGATCTCCGACGGCAGTGACGTCGCCGACACAAT
CGCTCACTTCTCAAGACGAGGCAACGCGGCGTTTGCCTTCTCAGCGGGACAGGCTCAGT
CGCTAACGTCACCCCTCCGCCAAGCCGCCGACCCAGGAGGTGTGGTCTCTCTCCAAGGCAG
GTTTGAAATCTTATCTTTAACCGGTGCTTTCTCTCCCTGGACCTTCCCACCCGGGTCAAC
CGGTTTAACGGTTTACTTTAGCCGGGGTCCAGGGTCAGGTCGTTGGAGGTAGCGTTGTAGG
CCCCTCTTAGCCATAGGGTCGGTCATGGTGATGTGCTGCTACTTTCTCTAACGCTACTTA
TGAGAGATTGCCCATGGAGAAGAGGAGACGGTGGCGGCTCAAGACAGATTACGGAGG
CGGTGACTCACCGCCAGAAATCGGTAGTAACCTGCCTGATCTATCAGGGATGGCCGGGCC
AGGCTACAATATGCCGCCGCATCTGATTCCAAATGGGGCTGGTCAGCTAGGGCACGAACC
ATATACATGGGTCCACGCAAGACCACCTTACTGACTCAGTGAGCCATTTCTATATATAAT
GGTCTATATAAATAAATATATAGATGAATATAAGCAAGCAATTTGAGGTAGTCTATTACA
AAGCTTTTGCTCTGGTTGGAAAAATAAATAAGTATCAAAGCTTTGTTGTTCTTAATGGA
AATATAGAGCTTGGGAAGGTAGAAAGAGACGACAT

>G1069 Amino Acid Sequence (domain in AA coordinates: 67-74)

MANPWWTNQSLAGMVDHSVSSGHHQNHQSLLLTKGDLGIAMNQSQDNDQDEEDDPREG
AVEVNRPRPRRPPGSKNPKAPIFVTRDSPNALRSHVLEISDGSVDADTIAHFSRRRQR
GVCVLSGTGSVANVFLRQAAAPGGVVSQGRFEILSLTGAFPLPGSPPPGSTGLTVYLAGV
QQQVVGGSVVGPLLAIGSVMVIAATFSNATYERLPMEEEDGGGSRQIHGGGDSPPRIGS
NLPDLGSMAGPGYNMPPHLIPNGAGQLGHEPYTWVHARPPY*

>G1070 (170..1144)

TCGACCAGCTTGGATTTGTTGTTTCATCATTACTACTCTCTTTCTTCTTCTAGCTAGCTA
GTTTTGACAGCAAAATAAGAACAAAAAAGGTCAACTAAAAAGATCTGTTCTTAGAT
CACTCTCTTCTTCTTTTCTTTTCTGATCCAATTCACCATTGAATCATAGATCATGGATCCAGT
ACAATCTCATGGATCACAAGCTCTCTACCTCCTCCTTTCCACGCAAGAGACTTTCAATT
ACATCTTCAACAACAGCAACAAGAGTTCTTCTCCACCATCACCAGCAACAAGAAACCA
AACCGATGGTGACCAACAAGGAGGATCAGGAGGAAACCGACAAATCAAGATGGATCGTGA

AGAGACAAGCGACAACATAGACAACATAGCTAACACAGCGGTAGTGAAGGTAAAGACAT
AGATATACACGGTGGTTTCAGGAGAAGGAGGTGGTGGCTCCGGAGGAGATCATCAGATGAC
AAGAAGACCAAGAGGAAGACCAGCGGGATCCAAGAACAAACCAAAACCACCGATTATCAT
CACACGGGACAGCGCAAAACGCGCTTAGAACCCACGTGATGGAGATCGGAGATGGCTGCGA
CTTAGTCGAAAGCGTTGCCACTTTTGCACGAAGACGCCAACGCGGCGTTTGCCTTATGAG
CGGTACTGGAAATGTTACTAACGTCCTACTATACGTCAGCCTGGATCTCATCCTTCTCCTGG
CTCGGTAGTTAGTCTTCACGGAAGGTTTCGAGATTCTATCTCTCTCAGGATCTTTCTCCC
TCCTCCGGCTCCTCTACAGCCACCGGATTGAGTGTTTACCTCGCTGGAGGACAAGGACA
GGTGGTTGGAGGAAGCGTAGTTGGTCCGTTGTTATGTGCTGGTCTGTCTGTCGTTGTCATGGC
TGCGTCTTTTAGCAATGCGGCGTACGAAAGGTTGCCCTTTAGAGGAAGATGAGATGCAGAC
GCCGGTTTCATGCGGAGGAGGAGGAGGATCATTGGAGTCGCCGCCAATGATGGGACAACA
ACTGCAACATCAGCAACAAGCTATGTCAGGTTCATCAAGGTTTACCACCTAATCTTCTTGG
TTCGGTTCAGTTGAGCAGCAACATGATCAGTCTTATTGGTCAACGGGACGACCAACCGTA
TTGATCAAATATACACACACTCATAATCGTTGCTAGCTAGCTAACGATGAATCATGAG
TTTAGTGGATATATATATGATTAAAAGAGGTTAGCTTATGAACATTAATAAGAGTTTGA
TTCTATCGAGCTTCATTATGTTTGGGTCATCGTTC

>G1070 Amino Acid Sequence (domain in AA coordinates: 98-120)
MDPVQSHGSQSSLPFFHARDQFLHLQQQQQEFFLHHHQQQRNQTDGDDQGGSGGNRQIK
MDREETSNDINIANNSGSEKIDIDIHGSGEGGGGGSGGDHQMTRRPRGRPAGSKNPKPK
PIIITRDSANALRTHVMEIGDGDLVESVATFARRRQRGVCVMSGTGNVTNVTIRQPGSH
PSPGVSVSLHGRFEILSLSGSFLPPAPPTATGLSVYLAGGQGVVGGSVVGPLLCAGPV
VMAASFNSAAYERLPLEEDEMOTPVHGGGGGSLSPPMMGQQLQHQQQAMSGHQLP
NLLGSVQLQQHHDQSYWSTGRPPY*

>G1076 (198..1076)

ATTTTAGTCTTCTTAACTTCTTCTCAATCCTCTCTCATATCTTTTTTCTTAGTTTAAA
TTTCAATAAAATAGAAAAAACATATACAAATCTACAGAGAAGAGAAGCTTTATTTTAAAT
CTTGTGTGTGTGTGTGTGTGTATATAATTTTATTTTTTTTCAAATTAAATCTCTTCT
TTGCTTTTGATGTGGGCATGGCTGGTCTTGATCTAGGCACAGCTTTTCGTTACGTTAATC
ACCAGCTCCATCGTCCCGATCTCCACCTTCACCACAATTCTCTCCGATGACGTCATC
CCGAGCCCGGATGGGTCAATTTACCGTCGACGAGCAAGACAACAACAACCATCAAG
GTCTTGACTTAGCCTCTGGTGGAGGATCAGGAAGCTCTGGAGGAGGAGGAGGTACGGCG
GGGGAGGAGACGTCGTTGGTTCGTCGTCACGTGGCAGACCACCGGATCCAAGAACAAC
CGAAACCTCCGGTAATTATCAGCGCGAGAGCGCAAACTCTAAGAGCTCACATTCTTG
AAGTAACAACCGCTGCGATGTTTTCGACTGCGTTGCGACTTATGCTCGTCGGAGACAGC
GAGGATCTGCGTTCGAGCGGTAGCGGAACGGTCACGAACGTCAGCATACGTCAGCCAT
CTGCGGCTGGAGCGGTTGTGACGCTACAAGGAACGTTGAGATTCTTCTCTCTCCGGAT
CGTTTCTCTCTCTCCGGCACCTCCCGGAGCAACGAGTTTGACAATTTTCTTAGCCGGAG
GACAAGGTGAGTGGTTGGAGGAAGCGTTGTGGTGAGCTTACGGCGGCTGGACCGGTGA
TTGTGATTGTCAGCTTCGTTTACTAATGTTGCTTATGAGAGACTTCCTTTAGAAGAAGATG
AGCAGCAGCAACAGCTTGGAGGAGGATCTAACGGCGGAGGTAATTTGTTTCCGGAGGTG
CAGCTGGAGGAGGAGGAGGACTTCCGTTCTTTAATTTACCGATGAATATGCAACCAAATG
TGCAACTTCCGGTGGAAGGTTGGCCCGGGAATTCGGGTGGAAGAGGTCCTTCTGATGTG
TATATATTGATAATCATTATATATATACCGGCGGAGAAGCTTTTCCGGCGAAGAATTTGC
GAGAGTGAAAGAAAGGTTAGAAAAGCTTTTAAATGGACTAATGAATTTCAAATTATCATCGT
GATTTCCGACATTGTCTTGTTCATCATGTTAAGCTTAGGTTTATTTTTTGTCTTTGTAG
AATTTTATGTTTGAATCCTTTTTTTTCTGTGAACTCTATTGTGTTCTGCTGCGAAGG
AAAAAAAATTCTCAAAAAAA

>G1076 Amino Acid Sequence (domain in AA coordinates: 82-89)
MAGLDLGTAFRYVNHQLHRPDLHLHNNSSDDVTPGAGMGHFTVDDNNDNNHQLDLAS
GGGSGSSGGGGHGGGDVVGRRPRGRPPGSKNPKPKPVIIITRESANTLRAHILEVTNGC
DVFDVATYARRRQRGICVLSSGTVTNVSIRQPSAAGAVVTLQGTFEILSLSGSFLPPP
APPGATSLTIFLAGGQGVVGGSVVVELTAAGPVIVIAASFNTVAYERLPLEEDEQQQL
GGGSGGGLPFVEAAGGGGLPFFNLPMNMQPNVQLPVEGWPGNSGGRGPF*

>G1089 (31..2427)

AAGTAAGAGAGCTTCTTAAGGAAGAAGAAGATGGGTTGTGCTCAATCAAAGATCGAGAAC
GAAGAAGCAGTTACTCGTTGCAAGAACGAAAACAATTGATGAAAGACGCCGTCACTGCT
CGTAACGCTTTCGCCGCCGCTCACTCAGCTTACGCTATGGCTCTTAAAAACACCGGAGCT

GCTCTTTCCGATTACTCTCACGGCGAGTTTTTAGTCTCTAATCACTCGTCTTCCTCCGCA
 GCTGCAGCAATCGCTTCTACTTCTCTCTCCCACTGCTATATCTCCTCCTCTTCCTTCT
 TCCACCGCTCCGGTTTCTAATTCAACCGCTTCTTCTCTCCGCTGCGGTTCCCTCAGCCG
 ATTCTGTACTCTTCTCTCTCTCTCTCCACCACCGCTTCTCTTCAACGTGCTGCT
 ACTATGCCGAGATGAACGGTAGATCCGGTGGTGGTCATGCTGGTAGTGGACTCAACGGA
 ATTGAAGAAGATGGAGCCCTAGATAACGATGATGATGACGATGATGATGATGATGACTCT
 GAAATGGAGAATCGTGATCGTTTGATTAGGAAATCGAGAAGCCGTGGAGGTAGTACTAGA
 GGAAATAGGACGACGATTGAAGATCATCATCTTCAGGAGGAGAAAGCTCCGCCACCTCCC
 CTTTGGCGAATTCCGCGCCAATTCCGCCGCCACGTGAGCATCAGCATCAACATCAGCAA
 CAGCAACAACAACCTTTCTACGATTACTTCTTCCCTAATGTTGAGAATATGCCTGGAAC
 ACTTTAGAAGATACTCTCCACAACCACAACCACAACCAAGGCTGTGCCTCCTCAA
 CCACATTCACCAGTCGTTACTGAGGATGACGAAGATGAGGAGGAGGAAGAGGAGGAAGAG
 GAGGAGGAAGAGGAGACGGTGATTGAACGGAAACCACTGGTGGAGGAAAGACCGAAGAGA
 GTAGAGGAAGTGACGATTGAATTGGAAGTTACTAATTTGAGAGGGATGAAGAAGAGT
 AAAGGGATAGGGATTCCCGGAGAGAGGAGAGGAATGCGAATGCCGGTGACTGCGACGCAT
 TTGGCGAATGTATTCAATTGAGCTTGATGATAATTTCTTGAAAGCTTCTGAAAGTGCTCAT
 GATGTTTCTAAGATGCTTGAAAGCTACTAGGCTCCATTACCATTCTAATTTTGAGATAAC
 CGAGGACATATTGATCACTCTGCTAGAGTGATGCGTGTAATTACATGGAATAGATCATT
 AGAGGAATACCAATGCTGATGATGGGAAAGATGATGTTGATTGGAAGAGAATGAACT
 CATGCTACTGTTCTTGACAAATTGCTAGCATGGGAAAAGAAGCTCTATGACGAAGTCAAG
 GCTGGCGAAGTCAAGAAATCGAGTACCAGAAAAGGTTGCTCATTTAAATCGGGTGAAG
 AAACGAGGTGGCCACTCGGATTCATTAGAGAGAGCTAAAGCAGCAGTAAGTCATTTGCAT
 ACAAGATATATAGTTGATATGCAATCCATGGACTCCACAGTTTCAGAAATCAATCGTCTT
 AGGGATGAACAACTATACCTAAAGCTCGTTACCTTGTTGAGGCGATGGGGAAGATGTGG
 GAAATGATGCAATACATCATCAAAGACAAGCTGAGATCTCAAAGGTGTGAGATCTCTA
 GATGTTTCAAGCGGTGAAAGAAACAAATGATCATCATCACGAACGCCACCATCCAGCTC
 TTGGCAGTGGTTCAAGAATGGCACACGCAGTTTTGCGAGGATGATAGATCATCAGAAAGAA
 TACATAAAAGCACTTGGCGGATGGCTAAAGCTAAATCTCATCCCTATCGAAAGCACACTC
 AAGGAGAAAGTATCTTCGCTCTCGAGTTCCCAATCCCGCAATCCAAAACTCCTCCAC
 GCTTGGTATGACCGTTTAGACAAAATCCCCGACGAAATGGCTAAAGTGCCATAATCAAT
 TTCGCAGCGGTTGTAAGCACGATAATGCAGCAGCAAGAAGACGAGATAAGTCTCAGAAAC
 AAATGCCAAGAGACAAGAAAAGAAATTTGGGAAGAAAATTAGACAGTTTGAGGATTGGTAC
 CACAAATACATCCAGAAGAGAGACCGGAGGGGATGAATCCGGATGAAGCGGATAACGAT
 CATAATGATGAGGTGCTGTGAGGCAATTCAATGTAGAACAAATTAAGAAGAGGTTGGAA
 GAAGAAGAAGAGCTTACCATAGACAAAGCCATCAAGTTAGAGAGAAGTCACTGGCTAGT
 CTTCGAATCGCCTCCCGAGCTTTTTCAGGCAATGTCGAGGTTGCGTATTCATGTTTCG
 GATATGTATAGAGCTATAACGTATGCGAGTAAGCGGCAAGCCAAAGCGAACGGCATCAG
 AAACCTAGCCAGGGACAGAGTTTCGTAAGAACTAATGTAAGATCAGAGTAATGTCTTCTTC
 TTCTTTGATCTTGAATATTTAAGCACACACATACATAACGTATAGCTAAATCTTTATC
 ATTGCTTTCTTATATTAAGGTTTGGCTTTTGTGAAGAGGTTTCTTACATATGAGATTCA
 TATAGTGTGATTCTTAAGGAACTGTTCTGTTGAGTAATAAGAAAGTTGTGTATTGAAA
 TAGAGTTGCATTGTTAATTTTG

>G1089 Amino Acid Sequence (domain in AA coordinates 425-500)
 MGCAQSKIENEEAVTRCKERKQLMKDAVTARNAFAAAHSAYAMALKNTGAALSDYSHGEF
 LVSNHSSSSAAAAIASTSSLPTAISPLPSSTAPVSNSTASSSSAAVPQPI PDLPPPPP
 PPPLPLQRAATMPENMNGRSGGGHAGSGLNGIEEDGALDNDDDDDDDSEMNDRDLIR
 KRSRGGSTRGNRTTIEDHHLQEEKAPPPPLANSRPIPPRQHQQHQQQQQQQPFYDYF
 FPNVENMPGTTLIEDFPQPQPQPTRPVPPQPHSPVVTEDDEDEEEEEEEEEEEETVIER
 KPLVEERPKRVEEVTTIELEKVTNLGMMKSKGIGIPGERRGMRMPVTATHLANVFIELDD
 NFLKASESAHDVSKMLEATRLHYHSNFADNRGHIDHSARVMRVITWNRSFRI PNADDGK
 DDVDLEENETHATVLDKLLAWKLYDEVKAGELMKIEYQKKVAHLNRVKKRGHSDSLE
 RAKAAVSHLHTRYIVDMQSMDSVSEINRLRDEQLYLKLVHLVEAMGKMWEMMQIHHRQ
 ABISKVLRLDVSQAVKETNDHHHERTIQLLAVVQEWHTQFCRMIDHQKEYIKALGGWLK
 LNLPIESTLKEKVSSPPRPVNPAPAIQKLLHAWYDRDLKIPDEMAKSAIINFAAVSTIMQ
 QQEDEISLRNKEETRKELGKIRQFEDWYHKYIQKRGPEGMNPDEADNDHNDVAVRQF
 NVEQIKKRLEEEEEEYHRQSHQVREKSLASLRTRLPQLFQAMSEVAYSCSDMYRAITYAS
 KRQSQSERHQKPSQGQSS*

>G1093 (1..531)

ATGGGTTATCCGGTGGGGTACACTGAGCTCCTCCTCCCAAGAATCTTCCTTCACTTACTC
TCTCTCTTAGGCTTAATACGAACACTCATAGACACGGGTTTTTCGGATATTGGGTCTACCC
GACTTTCTCGAATCCGACCCGGTTTCATCGTCATCGTCATGGCTGGAACCACCGTATATG
TCCACGGCGGCGCATCATCAACGAAAGCTCATTTTTCTTCCCAGTGGCGGCGAGGCTA
GCTGGAGAAATCTTGCCCGTCATCAGATTCTCGGAGCTAACTCGACCCGGATTCCGATCC
GGATCCGATTGCTGCGCGGTGTGCCTCCACGAGTTCGAGAACGATGACGAGATCCGACGG
CTGACGAATTGTCAACACATATTTACCCGGAGCTGTTTAGACCGTTGGATGATGGGTTAT
AATCAGATGACGTGTCCACTTTGTAGAACGCCGTTTATTTCTGATGAGTTACAAGTTGCT
TTTAACCAACGAGTTTGGTCTGAATCTGAACTTCTCGCAGAATCAAATTAG
>G1093 Amino Acid Sequence (domain in AA coordinates: 105-148)
MGYPVGYTELLLPRIFLHLLSLGLIRTLIDTGFRLGLPDFLES DPVSSSSSWLEPPYM
STAAHHHQESSFFFPVAARLAGEILPVIRFSELTRPGFGSGSDCCAVCLHEFENDEIRR
LTNCQHIFHRSCLDRWMMGYNQMTCLCRTPFISDELQVAFNQRVWSESELLAESN*

>G1127 (191..1351)

GACAGACTCTCTCTGTATGTGTGCGAGAAGCGAGAAGCGAGAGAGAGAGAGAGAGTTG
TTAGCTCACACGCTTTCTCTATTTTCTCGGAATTCACAAAACAGAAAGTTTCATCCTTTA
CGAGAATTAAGCCGAAAGAAACAATCTTTGAGTTTGATTCTTCTTCTTCTTCTCTCTCT
CTCTGCTCTAATGGATTCCAGAGACATCCACCGTCACATAACCAGCTTCAACCACCACC
GGGAATGTTAATGTCTCATTACCGTAACCCTAACGCCGCGCTTCAACCTTAATGGTTCC
CACTTCCACATCTCAACCGATTCAACACCTCGTCTTCTTTTGGCAATCAACAACAATC
TCAAACGTTTTCATCAGCAGCAACAACAACAATGGATCAGAAGACTCTTGAATCTCTTGG
ATTGGTGATGGATCACCTTCTTCTCAACCGATGCGATTGGGATCGATGATCAGAAATCA
GCAACTGCAAGTGAAGAAGAAGCGAGGAAGGCCGAGAAAGTATACTCCTGATGGTAGCAT
TGCTTTAGGTTTAGCTCCTACGTCTCTCTTCTCTGTCAGCTTCTAATTCTTACGGTGA
GGGTGGTGTGGAGATAGTGGTGAAATGGAACTCTGTTGATCCACCTGTTAAACGTAA
CAGAGGAAGGCCTCCTGGTTCTAGTAAGAAACAGCTTGATGCTTTAGGAGGAACCTCAGG
AGTTGGGTTTACACCTCATGTCAATTGAAGTGAACACAGGAGAGGACATAGCGTCAAAGGT
GATGGCTTTTTCGGATCAAGGGTCAAGAACAATTGTATTCTCTCTGCAAGTGGTGCGAT
TTCTAGAGTGATGCTTCGTCAAGCTTCTCATTCTAGTGAATCGTTACTTATGAGGGACG
ATTTGAGATCATTACTCTCTCAGGCTCAGTCTTGAATTATGAGGTAAATGGTTCCACCAA
CAGAAGTGGAACCTTGAGTGTGGCTTTGGCTGGACCTGATGGCGGCATCGTAGGTGGCAG
TGTAGTTGGTAATCTAGTAGCTGCAACACAAGTCCAGGTGATAGTGGGAAGCTTTGTTGC
AGAAGCAAAGAAACCGAAACAAAGTAGTGTAAACATTGCTCGGGGGCAGAATCCTGAACC
GGCTTCAGCGCCGGCTAACATGTTGAACCTTGGATCAGTCTCTCAAGGACCATCGAGCGA
GTCATCAGAAGAGAATGAGAGCGGTTCTCTGCAATGCACCGTGACAATAATAATGGGAT
ATATGGAGCTCAACAACAACAACAACAACCTCTCATCCTCATCAGATGCAAAATGTA
CCAACATCTTTGGTCTAATCATGGTCAATAAAATGAAGCGGAATTAATTTGTTTCCGTT
TTGGTTACGGTTATGGTTTGATTCTT
>G1127 Amino Acid Sequence (domain in AA coordinates:103-110, 155-162)
MDSRDIPPSHNQLQPPPGMLMSHYRNPNAASPLMVPTSTSQPIQHPRLPFGNQQSQT
HQQQQQQMDQKLTLES LGFGDGS PSSQPMRFGIDDQNLQVKKRGRPRKYTPDGSIALG
LAPTSPLLSAASNSYEGGVGDSGGNGNSVDPVVKRNRGRPPGSSKKQLDALGGTSGVGF
TPHVIEVNTGEDIASKVMFSDQGSRTICILSASGAVSRVMLRQASHSSGIVTYEGRFEI
ITLSGSLVLYEVNGSTNRSNLVALAGPDGGIVGGSVVGNLVAATQVQVIVGSFVAEAK
KPKQSSVNIARGQNPPEPASAPANMLNFGSVSQGPSSSESSEENESGSPAMHRDNNGIYGA
QQQQQQQLHPHQMQMYQHLWSNHGQ*

>G1131 (57..758)

TCGACTCCTCTCCTGATTGCTTCACCTTCTTCTTTACTACAGGTTTTAGCTCCTCAATGT
CCATGGATTGCTTAAGCTACTTCTTAACTACGATCCTCCTGTCCAGCTCCAGGATTGCT
TTATTCCTCGAGATGGATATGATTATCCCTGAAACCGATAGTTTCTTCTTCCAATCTCAAC
CGCAACTGGAGTTTCATCAGCCATTGTTTCAAGAAGAAGCTCCTTCACAGACCACTTTG
ACCTTTCTCGGACCACTTCTTCTCCTCGCAAGAAATCTTCTCCCTAACCTTAAACG
AAATCTTCAACGAAACACGACCTCGATTCTTCTTCTCCACGCAAAACGCCAGAGAC
TTGTTAACTCCAGCTACAATTGTAACTCAAACCATTTCCAGAGCCGTAACCCGAATT
TCTTCGACCTTTTCGGCGACACTGATTTGCTTCCAGAATCTGTACCTTCCAGGAGTTTC
GAGTTCCGGATTTCTCTTTAGCTTTCAAGGTAGGCCGGGAGATCAAGATGACTCAAAGA

AACCGACGCTTTTCATCTCAGAGCATCGCGGCTAGAGGGAGGAGAAGAAGAAATTGCAGAGA
AGACTCACGAGCTCGGAAAACTCATCCCCGGTGGCAATAAACTTAACACCGCCGAGATGT
TCCAAGCCGCCGCTAAGTATGTCAAGTTTTTGCAGAGTCAAGTTGGGATTCTCCAAGTGA
TGCAGACCACAAAGAAGGTAATAACCAACCCCAATAAGAAGCTTTATCATCCAATTGAAA
CTCTAATCGTGTTTTCTCACAGCTTCTTAATTTGTTTACGCAGGGTAGCTCTAATGTGC
AAATGGAAGCTCAGTATTTGCTTGAATCGCAAGCAATCCAGGAGAAGTTATCAACAGAGG
AAGTGTGTTTGGTACCGTGTGAAATGGTTCAAGATCTAACAACGAAGAAACCATTGCA
GAACCCGAATATTTCTCGAGAAATCAACAAGTTACTGTCTAAACATCTGGCTAACTAGT
TTTAGTTTCAAGCCTGAAGTTCTCTATGCCTAAATTTGTGTCTGTTATCGTTGTTTGTCT
TTCTTAGTTAGTGTGTTTGTCTTGTGATTTAGGGGCTAATTATCCTGGTTAATCTCCTCT
TAACTGGGAA

>G1131 Amino Acid Sequence (domain in AA coordinates: 173-220)
MSMDCLSYFFNYDPPVQLQDCFIPEMDMIIPETDSFFFQSQPQLFHFQPLFQEEAPSQTH
FDPFCDQFLSPQEIFLPNPKNEIFNETHDLDFLPTPKRQRLVNSSYNCTQNHQSRNP
NFFDPPGDTDFVPESCTFQEFVRPDFSLAFKVGGRGDQDDSKKPTLSSQSI AARGRRRRIA
EKTHELGKLI PGGNKLNTAEMFQAAAKYVKFLQSQVGILQLMQTTKKVITNPK*

>G1145 (243..1142)

GTGATTTCTCTCTGCCATTTCTTCGATTTGATTTCTGGGTTCTCTTCTCTCTCGTCTCTC
TTCTGCATGTTTCGCCACTCTACCTTAGAAAAAGGTTACTTTCCGCTCCGATTTAGGCT
CGATTTGATGAATTCGTCGTCGTGTGGCTATTTATCAAATTGAGCATTAGGGTTTCTGAT
TTGTGGGTTTCAAGATTGTTTTATCTATCTGTCTTGTGTTTTTGTCCGCTACAAAAGC
CTATGGATTCTCAGAGGGGTATTGTTGAACAAGCTAAATCTCAGTCTTGAATAGGCAAA
GCTCTCTTTACAGCTTAACACTTGATGAGGTTCAAATCACTGGGGAGTTCTGGTAAAG
CTCTGGGAAGCATGAACCTTGATGAGCTTTTGAAGAGTGTCTGTTCTGTTGAAGCTAATC
AGCCATCGTCTATGGCTGTCAATGGTGGAGCAGCTGCTCAGGAGGGTCTTCTCGCCAGG
GGAGTTTGACTTTGCCTCGGGATCTCAGCAAAAAGACTGTTGATGAGGTTTGGAAAGACA
TTCAGCAGAATAAGAATGGAGGTAGTGCTCATGAGAGGAGGGATAAGCAGCCTACACTTG
GGGAAATGACGCTTGAAGACCTGTTGTTGAAAGCAGGAGTGGTCACTGAGACTATCCCTG
GTTTCAACCATGATGGTCTCTGTTGGTGGTGGTAGTGCTGGTTCAGGTGCTGGTTTAGGGC
AAAACATTACTCAAGTTGGCCCATGGATTCAATATCATCAGCTCCCATCAATGCCACAGC
CTCAAGCATTTATGCCCTATCCGGTTTCAGATATGCAAGCAATGGTGTCTCAGTCTTCTT
TGATGGGTGGTTTGTGAGATACACAACTCCTGGAAGGAAGAGGGTAGCTTCAGGAGAAG
TTGTAGAGAAGACTGTAGAGAGGAGGCAGAAGAGAAAGATAAAGAACAGAGAGTCTGCTG
CTCGTTCCCGAGCTAGGAAACAGGCTTACACTCATGAGCTAGAGATCAAAGTTTACGGT
TAGAAGAAGAAAAACGAAAGACTCAGGAAGCAAAAGGAGGTGGAAAAATCCTCCCAAGTGT
ACCACCGCTGATCCCAAGCGGCAGCTCCGACGGACAAGCTCGGCTCCTTTCTGATCTCT
AAACTCTTTTGTCTTTTCTTTTCTTTTCTCTTCTGTGTCGGTTCACTTATAAAAAAGAGA
GGAAAACAGCTTTGTTTCTTTGTACATTCGCTAGACTTTCTTGACTTGGAGCAATTCTGT
TAACTTTAAATATTCTCGAGTTATTGTAGTAGCAGACTAGCAGCAGTAATGGTTTTTCAT
GAGTCCGATTGAAATTCAGAGATTGAACAGGAAAAAA

>G1145 Amino Acid Sequence (conserved domain in AA coordinates: 227-270)

MDSQRGIVEQAKSQSLNRQSSLYSLTLDEVQNLGSSGKALGSMNLDELLKSVCVSVEANQ
PSSMAVNGGAAAQEGLSRQGSLLPRDL SKKTVDEVWKDIQQNKNGGSAHERRDKQPTLG
EMTLEDLLLKAGVVTETIPGSNHDGPVGGGSAAGSAGLGQNTQVGPWIIQYHQLPSMPQP
QAFMPYPVSDMQAMVQSSSLMGGLSDTQTPGRKRVASGEVVEKTVERRQKRMINKRESAA
RSRARKQAYTHELEIKVSRLEENERLRKQKEVEKSSQVYHRLIPSGSSDGQARLLSDL*

>G1229 (123..1217)

TTTGGGCGGGTCTTTCTTTCCCTAAATCTTTCTTTTATTTTGCTGTTTTAAAAAAAATC
CAACCATAAGACAAAACAACGAACGAGGAAGAGAGAGAGAAGGATATATCTCTAATCA
CGATGCAGGAGATAATACCGGATTTTCTTGAAGAGTGTGAATTTGTGACACTTCACTAG
CCGGAGATGATCTATTTGCCATCTTAGAGAGTCTTGAAGGTGCCGGAGAGATATCTCCGA
CAGCTGCATCTACACCTAAAGATGGAACCAAGTTCCAAGGAGTTAGTTAAGGATCAAG
ATTATGAAAACCTCATCTCCTAAGAGGAAAAAGCAAAGACTAGAAACCAGGAAGAAGAGG
ACGAAGAAGAAAGACGGAGACGGAGAGCAAGCAAGAAGATAATAAGCAAGATGGGCAAC
AAAAGATGTCTCATGTAACCGTGAACGTAACCGGAGAAAGCAAATGAACGAGCACTTAA
CCGTTTTGCGTTCTCTTATGCCTTGTCTTCTACGTCAAACGGGGGGACCAAGCATCGATCA
TAGGAGGAGTTGTGGAGTACATAAGCGAGTTACAACAAGTTCTCCAATCTTTGGAAGCCA

AGAAACAACGTAAAACCTACGCCGAAGTCCTAAGCCCGAGAGTTGTCCCGAGCCCTCGTC
CTTCACCGCCTGTTCTAAGCCCAAGAAAACCGCCTCTTAGCCCGCGCATCAACCACCACC
AGATTACACCACCACCTACTTCTCCCTCCCATAGTCTCGAACACCTCAGCCCAACAAGCC
CATACCGGGCCATTCACCGCAACTACCACTCATCCCACAGCCTCCGCTTCGCTCTTACA
GCTCATTGGCCAGTTGCAGCAGCTTAGGAGATCCACCTCCATACTCTCTGCTTCATCTT
CTTCATCTCTTTCAGTTAGTAGTAACCATGAGAGTAGTGTGATCAATGAGCTTGTGCTA
ACTCAAAATCGGCTTTGGCTGATGTGGAAGTGAAGTTTTTCAGGAGCTAACGTGCTGCTCA
AAACGGTGTGCGATAAGATCCCGGGACAAGTTATGAAGATAATTGCTGCTCTTGAAGATT
TGGCTCTTGAGATTCTTCAGGTTAATATTAACACCGTCGACGAAACCATGCTTAATTCTT
TCACCATCAAGATTGAATTGAGTGCCAACTAAGTGCAGAAGAACTGGCTCAACAAATTC
AGCAAAACATTCTGCTAGTAAAGAAGGATTTAATATAGCTTCGTATAAACCTTAACGAGAG
AGCAGTACGTACTCACTTCTCTCTTAGTATCCCTTTAATTATCTTTTCAGTTTCTGTC
AAAGATATGGAGTTTAAAAAATAAAATTGTTATCTAAAGTTTAAATCAAATATTGATTA
ATTATAACTAATATAGGTATAAGTGAGTTTAAAGATTATCAGCTTCATAACAGCCATCG
TCATGTTTACTTTCTTTAAATTTTAGAATTTAGACGTACTCCTACCATGTAATTTTATT
TCTGTCTATTACATCAAGCATTGTAGCTGTAATTGCATATGAATGAACAATAGTGTATGAG
TGATCTCATGAATAATATCTTCTTGCAACACAAAAA

>G1229 Amino Acid Sequence (domain in AA coordinates: 102-160)
MQELIPDFLEECEFDVDSLGLDDLFALLESLEGAGEISPTAASTPKDGTSSKELVKDQD
YENSSPKRKKQRLERKEEEDGDEAEDNKQDQKMSHVTVERNRRKQMNHLT
VLRSLMPCFYVKRGDQASIIIGVVVEYISELQQVLQSLQLEAKKQKRTYAEVLSPRVPSRP
SPPVLSPRKPPLSPRINHHQIHHLHLLPPISPRTPQPTSPYRAIPPQLPLIPQPLRSYS
SLASCSSSLGDPPIYSPASSSSSPSVSSNHESVINELVANSKALADVEVKFSGANVLLK
TVSHKIPGQVMKIIAALEDLLEILQVNINTVDETMLNSFTIKIGIECQLSABEELAQQIQ
QTFC*

>G1246 (1..1746)
ATGATCATGTACGGAGGAGGAGGAGCAGGGAAGGACGGTGGATCCACCAATCACTTATCA
GACGGAGGAGTGATATTGAAGAAAGGTCCATGGACGGCGCGGAAGATGAGATACTTGCT
GCGTAGCTTAGAGAGAACGGTGAAGGGAATTGGAACGCCGTTTCAAAAAACACAGGTTTG
GCTCGTTGCGGCAAAAGCTGCCGTCTTCGATGGGCAATCACCTCCGACCAATCTGAAA
AAAGGCTCTTTCACCGGTGACGAAGAAGCTCTCATCATTACGCTTCATGCTCAGCTTGGT
AACAAATGGGCTCGCATGGCTGCTCAGTTACCGGGAAGAACAGACAACGAGATTAAGAAC
TATTGGAACACGAGATTGAAACGACTTCTTCGCCAAGGACTTCCTCTTATCCTCCAGAT
ATTATCCCTAACCATCACTCCATCCACATCCACATCATCAACAACAACAGCAACATAAC
CATCATCATCATCATCAACAACAACAACATCAACAAATGTATTTCAACCACAA
TCTTCACAACGAAACACACCATCATCTTCCCTCTTCCATCTCCAACACCAGCAACGCA
AAGTCTCATCATCTTCACTTTTCATACACGACTGCTAACCTCCTCCATCCACTTAGC
CCTCACACTCCAACAACACCATCTCAACTCTCTTCCACACCGCTCCACCACCACTTTCC
TCTCTTTTATGTTCCCTCGCAACAACCAATACCCGACCTTCCCTCTTGGCCTCCCG
CGTTCCCAATCAACAACAACAACGGAATTTCACTTTCCCTAGACCTCCACCTCTC
CTTCAACCGCTTCATCACTCTTCGCAAAACGTTACAACAATGCTAACACTCCTCTTAAT
TGCATCAACCGCTCTCAACCGCACCATTTTCCCTGTTTCAAGAGACTCCTACACTTCC
TTTCTTACATTGCCTTACCTTCCCAACCGCTCAAACCGCTACTTACCACAATACTAAT
AACCCTTACTCTTCTCTCTCTCTCTTCTTAAACCTTCTTCTTCTTCTTACCCTACA
TCAACTTCTTCCCCAAGCTTTCTTCACTCCCATACACTCCTTCTTCCACCTCATTTTAT
ACCAACCCAGTTTACTCCATGAAACAAGAGCAGCTCCCTTCAAACCAAATCCCCAAATA
GATGGCTTCAATAACGTCAACAACCTTACAGACAACGAGAGACAGAATCATAACCTTAAC
AGTTCCGGTGCTCATAGAAGAAGTAGTAGCTGCAGCCTCTTAGAGGATGCTTTCGAAGAG
GCCGAAGCTTTAGCCTCTGGAGGAGAGGCGGACCTCCAAAACGAAGACAACCTCACAGCT
TCTCTTCCGAACCAACAACAACCAACAACAACGACAACCTTCTTCTCGGTTAGTTTC
GGACATTATGATTCTTCTGACAACTTATGTTCTTGCAGATTGAAATCAAAGGAAGAA
GAGTCTCTTCAAATGAACACAATGCAGGAGGACATAGCTAAGCTTCTTGATTGGGGAGT
GATAGTGGAGAGATCTCTAATGGACAATCATCTGTTGTCACTGACGACAATCTTGTTCTT
GATGTTTATCAATTAGCTTCACTATTTCCGGCTGATTCTACAGCCGTCGTAGCCGCAACA
AACGACCAACACAACAAGAATAATAACAATAATTGTTCTGGGATGACATGCAGGGAATA
AGGTAG

>G1246 Amino Acid Sequence (domain in AA coordinates: 27-139)

MIMYGGGGAGKDDGGSTNHLSDGGVILKKGPWTAAEDEILAAYVRENGEGNWNNAVQKNTGL
ARCGKSCRLRWANHLRPNLKKGSFTGDEERLI IQLHAQLGNKWARMMAQLPGRTDNEIKN
YWNTRLKRLRLRQGLPLYPPDI IPNHQLHPHPHHQQQQQHNHHHHHHQQQQQHQQMYFQPO
SSQRNTPSSSPLSPPTPANAKSSSSFTFHTTTANLLHPLSPHTPNTPSQLSSSTPPPPPLS
SPLCSPRNQYPTLPLFALPRSQINNNNNNGNFTFPRPPPLLQPPSSSLFAKRYNNANTPLN
CINRVSTAPFSPVSRDSYTSFLTLPPYSPTAQTATYHNTNNPYSSSSPSFSLNPSSSSSYPT
STSSPSFLHSHYTPSSTSFHTNPVYSMKQEQLPNSQIPQIDGFNNVNNFTDNERQNHNLN
SSGAHRRSSSSCSLLEDVFEEAEALASGGGRPPKRRQLTASLPNHNNNTNNNDNFFSVSF
GHYDSSDNLCSLQDLKSKEEESLQMNMTMQEDIAKLLDWGSDSGEISNGQSSVVTDDNLVL
DVHQLASLFPADSTAVVAATNDQHNNNNNNCSWDDMQGIR*

>G1255 (138..1388)

CAGCTCAAACCTCTCTAGGACTACACTAAATCTAACTTTTTGCAGAGAGCAAAAGATTCAA
TAATTGAGATTGATCTCAAACCAAAGCTCTCGTGCTCTTGTCGTTGATGTTGGTTGTGT
AGACTTTGTATACATGATGAAAAAGTTTGGCGAATGCTGTTGGAGCGAAGACGCGCAGGG
CTTGCGACAGCTGCGTGAAGAGACGTGCACGGTGGTACTGCGCGGCCGACGATGCTTTTC
TTTGCCAGTCTTGCGACAGTTTGGTCCATTGAGCAAAACCTCTTGCTCGCCGCCACGAGA
GAGTCCGTTTGAAGACGGCTAGCCCGGCGGTCTGTAAGCATAGCAACCACTCATCAGCTT
CTCCTCCACATGAGGTGCGCCAGTGGCATCACGGGTTTACTCGTAAAGCTCGAACGCCAC
GTGGCTCTGGTAAGAAAAACAATTCGTCGATATTTTCATGACTTGGTTCCTGATATTAGTA
TTGAGGATCAGACAGACAACTATGAGCTTGAAGAGCAGCTGATCTGTCAAGTGCCGGTTC
TAGATCCGTTGGTGTCTGAGCAGTTCTTGAACGATGTGCTTGAGCCCAAGATCGAGTTTC
CTATGATCAGAAAGTGGTTTGTGATCGAGGAGGAGGAAGACAACGCTGAAAAGTTGTCTTA
ATGGATTTTCCCGACCGACATGGAGCTTGAGGAGTTTGTCTGCTGACGTGGAGACTCTGC
TCGGTCGCGGGTTAGACACGGAGTCTGATGCCATGGAGGAGCTAGGGTTATCTAATTCAG
AGATGTTCAAATCGAAAAAGATGAGATTGAAGAAGAAGTAGAAGAGATAAAAGCCATGA
GCATGGATATATTTGATGATGATCGAAAAGACGTGGATGGAACAGTACCGTTTGAGCTAA
GC'TTGATTACGAGTCGTACACAAGACGTCCGAAGAAGAGGTAATGAAGAACGTTGAAA
GTAGTGGTGAATGTGTTGTTAAGGTGAAAGAGGAAGAACATAAGAATGTTCTGATGCTAA
GATTAACTATGACTCGGTGATATCCACTTGGGGAGGTCAAGGTCCACCGTGGAGTTTCAG
GAGAGCCACCGGAACGAGACATGGACATCAGCGGTTGGCCAGCCTTTTCCATGGTGGAGA
ATGGAGGAGAAAAGTACTCATCAGAAAGCAATACGTTGGTGGATGTTTACCATCAAGTGGGT
TTGGAGATGGAGGTAGAGAAGCTAGAGTTTCGAGATACAGAGAGAAGAGGAGGACAAGGT
TGTTTTCTAAGAAGATACGGTACGAGGTACGTAAATGGAATGCAGAGAAAAGACCACGAA
TGAAAGGAAGATTTCGTGAAGAGAGCCTCGCTCGCTGCTGCTTACCATTAGGTGTTA
ATTACTGAATAGTTAATATCTATTGATGTTATATCTCACTTTACAAATTTCCGTGAATCT
TTTTTCTTCTGAAACAACAGAAGTTATTTTGGCACTTAATTGTGCTTTGAGGACTTGTAT
GTACATAGAAGTAACCAATAATAATGTGACTTTTACTA

>G1255 Amino Acid Sequence (domain in aa coordinates: 18-56)

MKSLANAVGAKTARACDSCVKRRARWYCAADDAFLCQSCDSLHVSANPLARRHERVRLKT
ASPAVKHNSNHSSASPPHEVATWHHGFTRKARTPRGSGKKNSSIFHDLVPDISIEDQTD
NYELEEQILICQVPVLDPLVSEQFLNDVVEPKIEFPMIRSGLMIEEEDNAESCLNGFFPT
DMELEEFADVETLLGRGLDTESYAMEELGLSNSEMFKIEKDEIEEEVEEIKAMSMDIFD
DDRKDVDTGVPFELSFDYESSHKTSEEEVMKNVSSGECVVVKVEEHNKVLMLRLNYDS
VISTWGGQGPWSSGEPPERDMDISGWPAFSMVENGGESTHQKQYVGGCLPSSGFGDGGR
EARVSRYREKRRTRLFSKKIRYEVRLNAEKRPMPKGRFVKRASLAAAASPLGVNY*

>G1304 (1..978)

ATGGGGCGATCACCATGTTGCGATGAGAATGGTCTAAAGAAAGGGCCATGGACACAAGAG
GAGGATGATAAACTGATAGATCACATTCAAAAACATGGCCATGGCAGCTGGAGAGCTCTT
CCAAAGCAAGCCGTTTAAACCGATGCGGAAAGAGTTGTAGATTAGATGGACCAACTAC
TTGAGACCTGACATCAAGAGAGGAAATTTCACTGAAGAGGAAGAACAACCTATTATCAAC
CTCCATTCCTTCTTGGAACAAGTGGTCTGTCGATAGCCGGTAATCTTCTGGAAGAACG
GACAATGAAATAAAAACTATTGGAACACACATTTGAGAAAGAACTTCTCCAATGGGG
ATTGATCCGGTGACCCATAGGCCAAGAACCAGCATCTAAACGTTTTAGCAGCTCTCCCG
CAGCTTATAGCCGCGCAAAATTTCAACAGCCTCTTGAATCTCAACCAAAATGTGCAACTG
GATGCAACAACCTCTTGCTAAAGCTCAACTGCTACACACTATGATTCAAGTCCTTAGCACC
AATAACAACACCACCAATCTTCTTTTCTTCATCAACTATGCAAAACAGTAACACCAAT
CTCTTTGGCCAAGCTTCTTACTTAGAGAACCAAAATCTTTTGGTCACTCTCAAACTTC

TCTCACATTCTTGAGGATGAGAATTTGATGGTCAAAACCCAAATTATTGATAACCCTTTG
GACTCTTTTTCTTCCCCCATAACAACCCGGTTTTCAAGATGATCATAATCACTCCCTCTA
TTGGTTCGGGCGTCTCCTGAAGAATCTAAAGAACTCAAAGGATGATCAAGAACAAGAC
ATCGTCGATTACCATCATCATGATGCTTCAAACCCCTTCATCATCAAACTCAACGTTTACA
CAAGATCATCATCACCCATGGTGTGACACTATTGATGATGGAGCAAGTGATTCTTTTTGG
AAAGAGATAATAGAGTAA

>G1304 Amino Acid Sequence (conserved domain in AA coordinates:13-118)

MGRSPCCDENGLKKGPWTQEEDDKLIDHIQKHGHSWRALPKQAGLNRCGKSCRLRWNTNY
LRPDIKRGNFTEEEEQTIINLHSLLGKWSIAGNLPGRITDNEIKNYWNTHLRKKLLQMG
IDPVTHRPRTDHLNVLAALPQLIAAANFNSLLNLQNVQLDATTAKAQLLHTMIQVLST
NNNTNPSFSSSTMQNSNTNLFQASYLENQNLFQSQNFSHILEDENLMVKQTQIDNPL
DSFSSPIQPGFQDDHNSLPLLVSPASPEESKETQRMINKNDIVDYHHHDASNPSSSNSTFT
QDHHHPWCDTIDDGASDSFWKEIIE*

>G1318 (7..849)

AAAAATATGAGGAAGCCAGAGGTAGCCATTGCAGCTAGTACTACCAAGTAAAGAAGATG
AAGAAGGGACTTTGGTCTCCTGAGGAAGACTCAAAGCTGATGCAATACATGTTAAGCAAT
GGACAAGGATGTTGGAGTGATGTTGCGAAAAACGCAGGACTTCAAAGATGTGGCAAAAGC
TGCCGTCTTCGTTGGATCAACTATCTTCGTCCTGACCTCAAGCGTGGCGCTTCTCTCCT
CAAGAAGAGGATCTCATCATTCGCTTTCATTCATCCTCGGCAACAGGTGGTCTCAGATT
GCAGCACGATTGCCTGGTCCGACCGATAACGAGATCAAGAATTTCTGGAACCTCAACAATA
AAGAAAAGGCTAAAGAAGATGTCCGATACCTCCAACCTAATCAACAACCTCATCTCATCA
CCCAACACAGCAAGCGATTCTCTTCTAATTCGCATCTTCTTTGGATATTAAAGACATT
ATAGGAAGCTTCATGTCCTTACAAGAACAAGGCTTCGTCAACCCTTCCTTGACCCACATA
CAAACCAACAATCCATTTCCAACGGGAAACATGATCAGCCACCCGTGCAATGACGATTTT
ACCCCTTATGTAGATGGTATCTATGGAGTAAACGCAGGGGTACAGGGGAACTCTACTTC
CCACCTTTGGAATGTGAAGAAGGTGATTGGTACAATGCAATATAAAACAACCACTTAGAC
GAGTTGAACACTAATGGATCCGGAACGCACCTGAGGGTATGAGACCAGTGAAGAATTT
TGGGACCTTGACCAGTTGATGAACACTGAGGTTCTTCTGTTTTACTTCAACTTCAAACAA
AGCATATGAATATTTTTACGTCATCTTATTCTTTTTTCTATTGCGGTTTATACTCAAGAT
TCTTAGCCACACACACATAAATGCAAAATATATATACATTGTTAGAGAGTATTTTGTATTT
CGTATAATCTTTTTCGTACTAGGGCTTGAGCCTTGAGGTCCCATGTAACGATTAGTCAATG
TAAACATATATCTATAATAATAATAATAAAGAAATAATAAGCACATAAAAAAAAAAAAA
A

>G1318 Amino Acid Sequence (domain in AA coordinates: 20-123)

MRKPEVAIAASTHQVKMKKGLWSPEEDSKLMQYMLNSNGQGCWSDVAKNAGLQRCGKSCR
LRWINYLRLPKRGAFSPQEDLIIRFHSILGNRWSQIAARLPGRITDNEIKNFWNSTIKK
RLKKMSDTNLIINNSSSSPNTASDSSSNSASSLDIKDIIGSFMSLQEQGFVNPSLTHIQ
NPFPTGNMISHPCNDDFTPYVDGIYGVNAGVQGEYFPFPLECEEGDWYNNINHLDEL
NINGSGNAPEGMRPVVEEFWDLDQLMNTTEVPSFYFNFQSI*

>G1320 (39..788)

GAAGATCATAAAGATCAAAAGGAGAGAGGTATTAATAAATGATGTGTAGTCGAGGCCATT
GGAGACCTGCAGAAGACGAGAAGCTAAGAGAACTCGTCGAGCAATTTGGTCCTCATAATT
GGAACGCCATAGCTCAGAAGCTCTCTGGTTCGATCTGGTAAGAGTTGTAGATTGAGATGGT
TTAATCAATTGGATCCTAGGATTAACCGAAACCCTTTCACGGAGGAAGAAGAAGAAAGGC
TTTTAGCGCCTCATCGGATCCATGGGAACAGATGGTCTGTGATCGCTAGATTTTTCCCG
GTCGAACTGATAACGCTGTTAAAAACCATTTGGCACGTCATCATGGCTCGTGGCCGAG
AACGGTCCAAGCTCCGTCCACGAGGCCTTGCCATGATGGCACGGTGGCTGCGACTGGGA
TGATTGGTAATTATAAAGACTGCGATAAGGAGAGAAGATTGGCAACCACAACCGCTATCA
ATTTTCCTTATCAATTCTCTCATATTAATCATTTTCAAGTCCTCAAAGAGTCCTTGACCG
GAAAGATCGGGTTCAGAAATAGTACTACTCCAATACAAGAAGGAGCAATAGACCAAACCTA
AACGACCGATGGAGTTCTACAATTTTCTCCAAGTAAACACGGATTGGAAGATACACGAAT
TGATAGATAATTCAAGAAAAGACGAAGAAGAAGATGTCGATCAAAACAACCGAATTCGTA
ACGAGAATTGTGTTCCATTTTTCGACTTTTGTCTGTGTTGGAAGTCTGCCTCTCAGGGTT
TATGTTAATTTGTCCGTACCACTACTATAAGGTGGACCATATGTTAACTAAAGATAA
TGTAAGAAAGTACTAATCAATTAGAGCTCCTGTTTGAGCCAAATGTGAAATTAGTTAAGA
CATCCCAAACATTTTCTGTATAACACATATAAGGTTGTACTTTTATCAGGTCTAATTTT
CTATTTTTTATTTAAGGATGTTTAATCAGACCCATAACCATTCGATAAAAAAAAAAAAAA

>G1320 Amino Acid Sequence (domain in AA coordinates: 5-108)
MMCSRGHWRPAEDEKLRELVEQFGPHNWNIAIAQKLSGRSGKSCRLRWFNQLDPRINRNPF
TEEEEEERLLAPHRIHGNRWSVIARFFPGRTDNAVKNHWHVIMARRGRERSKLRPRGLGHD
GTVAATGMIGNYKCDKERRRLATTTAINFPYQFQSHINHFQVLKESLTGKIGFRNSTTPIQ
EGAIDQTKRPMEFYNFLQVNTDSKIHLEIDNSRKDEEEDVDQNNRIRNENCVPFFDFLSV
GNSASQGLC*

>G1330 (36..959)
GTACCGGCGACCTCTTTGTGGGTCACTCTTCATCAATGGGTGACAAAGGAAGGAGCTTAA
AGATCAACAAGAACATGGAGGAATTCACGAAAGTGAAGAAGAAATGGACGTAAGGAGAG
GTCCATGGACAGTTGAGGAAGATTTAGAGCTCATCAATTACATTGCTAGTCATGGTGAAG
GTCGATGGAACCTCTCTCGCTCGTTGCGCCGAACTCAAAGGACCGGAAAAAGCTGCAGAC
TTCGGTGGCTGAACCTATCTCCGACCAGATGTGCGCCGTGGAACATAACCCCTCGAAGAAC
AACTCTTGATTCTTTGAACCTCACACACGTTGGGGCAATAGATGGTCTAAGATTGCACAAT
ATTTACCAGGAAGAACGGATAACGAGATCAAAAACCTATTGGAGAACACGTGTTCAAAGC
ATGCAAAACAGCTTAAATGCGACGTGAACAGTCAACAATTTAAAGACACCATGAAGTATC
TTTGGATGCCTCGGCTCGTAGAAAGGATCCAAGCCGCTCCATCGGGTCTGTTTCCATGT
CATCTTGCGTCAACACCTCCTCAGATCAGTTCGTGATCAACAACAACAACCAACAACG
TGGATAATTTGGCTTTAATGAGTAACCCCTAATGGTTACATCACGCCGGATAATTCCAGCG
TGGCAGTATCTCCTGTATCAGATTTGACGGAGTGTCAAGTGAGTAGTGAAGTGTGGAAGA
TTGGTCAGGATGAGAATTTGGTGGATCCAAAAATGACATCGCCGAATTATATGGATAATA
GCAGTGGACTATTAAACGGAGATTTTACGAAGATGCAAGATCAAAGTGACCTTAATTGGT
TTGAAAATATTAATGGGATGGTACCAAATTATTCGGACAGTTTTTGGAAACATTGGAATG
ATGAAGACTTCTGGCTCTTACAACAACATCAACAAGTCCACGACAATGGAAGCTTCTGAA
TAGACAAGAAGCTATGCGGCC

>G1330 Amino Acid Sequence (domain in AA coordinates: 28-134)
MGDKGRSLKINKNMEEFKVEEEMDVRRGPWTVEEDLELINYIASHGEGRWNSLARCAEL
KRTGKSCRLRWLNLYLRPDVRRGNITLEEQLLILELHTRWGNRWSKIAQYLPGRDNEIKN
YWRTRVQKHAKQLKCDVNSQQFKDTMKYLWMPRLVERIQAASIGSVSMSSCVTTSSDQFV
INNNNTNNVDNLALMSNPNGYITPDNSSVAVSPVSDLTECQVSSEVWKIGQDENLVDPKM
TSPNYMDNSSGLLNGDFTKMQDQSDLNWFENINGMVPNYSDFSFWNIGNDEDFWLLQHQHQ
VHDNGSF*

>G1352 (79..900)
GCGCGATTAAAACTCTCAACTTTTCTCTCAAATTTCTGATCCTTTGATCCAACAGTTAG
AAGAAGATTCATCTGATCATGGCCCTCGAAGCGATGAACACTCCAACCTCTTCTTTCACC
AGAATCGAAACGAAAGAAGATTTGATGAACGACGCCGTTTTCATTGAGCCGTGGCTTAAA
CGCAAACGCTCCAAACGTCAGCGTTCTCACAGCCCTTCTCGTCTTCTTCTCACCGCCT
CGATCTCGACCCAAATCCCAAGATCAAGATCTTACGGAAGAAGAGTATCTCGCTCTTTGT
CTCCTCATGCTCGCTAAAGATCAACCGTCGCAAACGCGATTTTCATCAACAGTCGCAATCG
TTAACGCCGCCGCCAGAATCAAAGAACCTTCCGTACAAGTGTAACTGTGTAAAAAGCG
TTTCTTCTCTATCAGGCTTTAGGCGGTCACAAAGCAAGTCACCGAATCAAACCACCAACC
GTAATCTCAACAACCGCCGATGATTCACAGCTCCGACCATCTCCATCGTCGCCGGAGAA
AAACATCCGATTGCTGCCTCCGGAAGATCCACGAGTGTTCAATCTGTCAAAAGTGTTT
CCGACGGGTCAAGCTTTAGGCGGTCACAAACGTTGTCACTACGAAGGCAACCTCGGCGGC
GGAGGAGGAGGAGGAAGCAAATCAATCAGTCACAGTGAAGCGTGTGAGCACGGTATCG
GAAGAAAGGAGCCACCGTGGATTTCATCGATCTAAACCTACCGGCGTTACCTGAACTCAGC
CTTCATCACAATCCAATCGTCGACGAAGAGATCTTGAGTCCGTTGACCGGTAAAAAACCG
CTTTTGTGACCGATCAGCACCAAGTCATCAAGAAAGAAGATTTATCTTTAAAAATCTAA
TACTCGACTATTAATCTTGTGTGATTTTTTTTCGTTACAACCATAGTTTCATTTTCATTT
TTTTAGTTACAAATTTTAAATTGTTCTGATTTGGATTGAAA

>G1352 Amino Acid Sequence (domain in AA coordinates: 108-129,167-188)
MALEAMNTPTSFTRIETKEDLMNDVFIPLWLRKRKRQRSHSPSSSSSPPRSRPKS
QNQDLTEEEYLALCLLMLAKDQPSQTRFHQQSQSLTPPESKNLPYKCNVCEKAFPSYQA
LGGLHAKSHRIKPPTVISTTADDSAPTISIVAGEKHPHIAASGKIHECSIHKVFTPTGQAL
GGHKRCHYEGNLGGGGGGGSKSISHSGSVSSTVSEERSHRGFIDLNLPALPELSLHNPPI
VDEEILSPLTGKPLLLTDHDQVIKKEDLSLKI*

>G1354 (1..1047)
ATGGAAAGTCTCGCACACATTCCTCCCGGTTATCGATTCCATCCGACCGATGAAGAACTC

GTTGACTATTATCTCAAGAACAAAGTTGCATTCCCGGGAATGCAAGTTGATGTTATCAAA
GATGTTGATCTCTACAAAATCGAGCCATGGGACATCCAAGAGTTATGTGGAAGAGGGACA
GGAGAAGAGAGGGAATGGTATTTCTTTAGCCACAAGGACAAGAAATATCCAACCTGGGACA
CGAACCAATAGAGCAACGGGCTCCGGATTTTGGAAAGCAACGGGTCGAGACAAGGCCATT
TACTCAAAGCAAGAGCTTGTGGGATGAGGAAGACTCTTGTCCTTTTACAAAGGTAGGGCC
CCAAATGGTCAGAAATCTGATTGGATAATGCACGAATACCGTCTTGAGACCGATGAAAAT
GGACCGCTCATGAGGAAGGATGGGTGGTTTGTGCGCTTTCAAGAAGAAGCTAACCACG
ATGAACCTACAACAATCCAAGAACAATGATGGGATCATCATCAGGCCAAGAATCTAAGTGG
TTCACGCAGCAAATGGATGTGGGGAATGGTAATTACTATCATCTTCTGATCTAGAGAGT
CCGAGAATGTTTCAAGGCTCATCATCATCATCACTATCATCATTACATCAGAATGATCAA
GACCCCTTATGGTGTGCTACTCAGCACTATTAACGCAACCCCACTACAATAATGCAACGA
GATGATGGTCATGTGATTACCAATGATGATGATCATATGATCATGATGAACACAAGTACT
GGTGATCATCATCAATCAGGATTACTAGTCAATGATGATCATAATGATCAAGTAATGGAT
TGGCAAACGCTTGACAAGTTTGTGCTTCTCAGCTAATCATGAGCCAAGAAGAGGAAGAA
GTTAACAAAGATCCCATCAGATAATTCTTCGAATGAAACATTTTCATCATCTCTCTGAAGAG
CAAGCTGCAACAATGGTTTCGATGAATGCTTCTTCTCTCTCTCTCCATGTTCTCTCTAC
TCTTGGGCTCAAAATACACACACGTAA

>G1354 Amino Acid Sequence (domain in AA coordinates: TBD)

MESLAHIPPGRYRHPTEDELVDYLLKKNKVAFFPMQVDVIKDVLDYKIEPWDIQLCGRGT
GEEREWYFFSHDKKYPTGTRTNRATGSGFWKATGRDKAIYSKQELVGMKRLTVFYKGRA
PNGQKSDWIMHEYRLTDENGPPHEGWVVCRAFKKKLTMMNYPNPTMMGSSSGQESNW
FTQQMDVGNNGNYHLPLDLESPRMFQSSSSSLSHQNDQDPYGVVLSTINATPTTIMQR
DDGHVITNDDHIMMNTSTGDHHSGLLVNDDHNDQVMDWQTLDKFVASQLIMSQEEEE
VNKDPDSNSSNETFHLSEEQAATMVMNASSSSSPCSFYSWAQNTHT*

>G1360 (1..1257)

ATGGGAGATAGAAACAACGACGGTGATCAGAAAATGGAGGATGTATTGTTGCCCGGATTT
AGGTTTTATCCAACCGACGAAGAGCTCGTAAGCTTCTACCTGAAGCGGAAGTTCAACAC
AACCTCTCTCCATTGAGCTCATAAGACAACTCGATATCTACAAATATGACCCCTGGGAT
CTTCCAAAGTTTGCATGACGGGTGAAAAGAATGGTACTTTTATGTCCAAGGGACAGG
AAGTATAGGAACAGCTCGAGGCCAAACCGAGTGACCGGAGCTGGTTTTTGGAAAGCCACG
GGAACGGACCGGCCGATATACTCGTCAGAAGGAAACAAATGCATAGGTTTAAAGAAGTCC
TTAGTGTTCTACAAAGGAAGAGCAGCGAAAGGAGTTAAGACTGATTGGATGATGCATGAG
TTTCGTTTGCCTTCTCTCTCCGAACCATCTCTCTCTTAAGAGATCTTTCGACTCTCTCT
GTCTCTCCCAACGATTCATGGGCTATATGCAGAATCTTCAAAAAGACCAACACAACGACC
CTAAGAGCTCTCTCTCACTCTTTTGTTCCTCGTTACCACCAGAAACAAGCACCAGACACA
ATGTCTAACAAAAGCAATCAAAACATACCATTTTTCTTCAGACAAGATCCTCAAACCT
AGCTCTCACTTCCAGTTTACCATGAGAATATGAACACTCCAAAACCTAGTAATAGTACA
ACTCCATCCGTTTCCACTATAAGTCCCTTCTCTTACTTGGATTTCACTTCATACGACAAA
CCCACCAACGTTTTCAATCCGGTTTCATGTTTAGACCAACAATACCTCACAATCTCTTT
CTTGCCACACAAGAAACACAACCTCAGTTTCCAGGCTCCCTCGTCAAATGAAATCCCA
TCGTTTCTGCTAAACACGTCTTCAGATTGACCTTCTTGGGAGAATTACAGAGCCATATC
GACCTCAGCGCAGTGTGGCCCAAGAGCAATGTCCCCGCTTGTAAAGCTACCACAGGAG
TATCAAGAGACGGGATTCGAAGGAAATGGTATAATGAAGAACATGCGTGGTTCCAATGAA
GATCATCTTGGTGATCATTGCGACACACTTCGGTTTGTATGATTTCACCTCAACAATTAAT
GAGAACCATCGTCATCATCAAGACCTGAAACAGAACATGACATTGCTGGAGAGTTATTAT
TCTTCTTTATCGTCCATCAATAGCGATTTGCCAGCTTGTCTTCTCCAGTACAACCTGA

>G1360 Amino Acid Sequence (conserved domain in AA coordinates: 18-174)

MGDRNNDGDQKMEDVLLPGFRHPTDEELVSFYLRKRVQHNPLSIELIRQLDIKYDPWD
LPKFAMTGEKEWYFYCPRDRKYRNSRPNRVTGAGFWKATGTDRIYSSEGNKICIGLKKS
LVFYKGRAAGVKTDWMMHEFRPLSLSEPPSKRFFDSPVSPNDSWAICRIFKKINTTT
LRALSHSFVSSLPPETSTDMSNQKQSNITYHFSDDKILKPSSHQFHHENMNTPKTSNST
TPSVPTISPFYLDFTSYDKPTNVFNPVSCLDQQLTNLFLATQETQPQFRLPSSNEIP
SFLNLTSSDSTFLGEFTSHIDLSAVLAQEQCPPLVSLPQEQETGFEGNGIMKNMRGSNE
DHLGDHCDTLRFDDFTSTINENHRHQDLKQNMTLLESYSSSLSSINSDLPACFSSTT*

>G1364 (1..537)

ATGGCGGAGTCGACGGCCAAGAGTCCCGGAGGCTGTGGAAGCCATGAGAGTGGTGAGAT
CAAAGTCCAGGTCGTTACATGTTCTGTGAGCAAGATAGGTTTCTTCCGATTGCTAACATA

AGCCGTATCATGAAAAGAGGTC'TTCCTGCTAATGGGAAAATCGCTAAAGATGCTAAGGAG
ATTGTGCAGGAATGTGTCTCTGAATTCATCAGTTTTCGTACCAGCGAAGCGAGTGATAAA
TGTCAAAGAGAGAAAAGGAAGACTATTAATGGAGATGATTGCTTTGGGCAATGGCTACT
TTAGGATTTGAAGACTACATGGAACCTCTCAAGGTTTACCTGATGAGATATAGAGAGGGT
GACACAAAGGGATCAGCAAAAGGTGGGGATCCAAATGCAAAGAAAGATGGGCAATCAAGC
CAAAATGGCCAGTTCTCGCAGCTTGCTCACCAAGGTCCTTATGGGAACTCTCAAGTAACT
TTTCCTCTCTTCTCTTCACTCAAGCAATACGCATCATTCTCTTCTAATTTGTTAA
>G1364 Amino Acid Sequence (conserved domain in AA coordinates: 29-120)
MAESQAKSPGGCGSHESGGDQSPRSLHVREQDRFLPIANISRIMKRGLPANGKIAKDAKE
IVQECVSEFISFVTSEASDKCQREKRKTINGDDLLWAMATLGFEDYMEPLKVYLMRYREG
DTKGSAGKGDPNAKDKGQSSQNGQFSQLAHQGPYGN SQVTFPLFSSHSSNTHHSLLIC*
>G1379 (68..622)
CTCTGCCTCTCTCTCTCTCAAAACCCATCTCGAAAGTCTTCTCTTTTCGAGGGTTTAG
ATCTTCCATGGAAGGCGCGGAGTTGCTGACGTGGCTGTCCCCGTACGAGGAAGAGAGA
CAGACCTTACAAAGGAATTAGGATGAGGAAGTGGGGAAAGTGGGTGGCGGAGATTCGTGA
GCCTAAACAAGCGCTCTAGGTTATGGCTTGGCTCTTACTCTACTCCCGAGGCGGCGCGCG
AGCTTACGACACGGCGGTTTCTATCTTAGAGGACCTACGGCGAGGCTTAACCTCCCTGA
GCTTCTTCTGGGGAGAAATTCTCCGACGAGGATATGTCGGCTGCGACCATCAGGAAGAA
AGCCACGGAGGTCCGTGCTCAGGTTGATGCTTGGGCACGGCGGTGCAAAATAACGCCA
CCGTGTTTTTGGTCAGAATCGAGATAGTGATGTGGATAATAAGAATTTTCATCGGAATTA
TCAAACCGGTGAACGAGAAGAAGAAGAAGATGAGGATGACAAGAGATTGAGGAGTGG
CGGCCGGTTATTGGATCGGGTTGACTTGAATAAATTACCCGACCCGAAAGCTCCGATGA
AGAATGGGAAAGCAAACATTAAAAATATATAGTTTGGAGCGGTGGCTGTTGCTAACGTAC
GCCAACGGCTTGCTTCTACGAATCATTAGCGCGGTTTATGATTTTTTTTTTTTTTTTTT
CATTATCTGAAAATTTAGGGCTTTTTAGTTATTAATTTTTGTTTTGTTTTTTTCTTTCT
TGCGAGTTTTGCGGTTTTATGGAATTTTAGGCTATTGCTTAACGAAAAAAAAAAAAAAAA
>G1379 Amino Acid Sequence (domain in AA coordinates: 18-85)
MEGGGVADVAVPGTRKRD RPYK GIRM RKWGWV AEIREPNKRSRLWLGSYSTPEAAARAY
DTAVFYLRGPTARLNFPELLPGEKFSD E DMSAATIRKKATEVGAQVDALGTAVQNNRHRV
FGQNRDSDVDNKNFHRNYQNGEREBEEEEDEDDKRLRSGRLLDRVDLNLKLPDPESSEDEW
ESKH*
>G1384 (33..977)
GTACATTTTTTTTTGTATTTTCAGGAAACTCCGATGGCGGATCTCTTCGGTGGTGGCCACG
GCGCGAGCTTATGGAAGCACTTCAACCTTTTTACAAAAGTGCTTCCACGTCTGCTTCAA
ATCCTGCGTTTGCGTCTCAAACGATGCGTTTGCGTCTGCCCCAAACGACCTATTTTCTT
CTTCTTCTTACTATAATCCTCATGCATCTTTATTCCTTTCACATTCACAACCTCTTACC
CGGATATTTATTCTGGATCCATGACCTATCCATCTTCATTTCGGGTGCGATCTTCAACAAC
CCGAAACTACCAATCTCAGTTCCATTACCAAAACACTATCACTTACACTCACCAAGACA
ACAACACTTG CATGCTTAACCTT CATTGAGCCGAGCCAACCGGGTTTTATGACCCAACCGG
GTCCGAGTTTCGGGTTTCGGTTTCAAAACCGGCTAAGCTCTATAGAGGAGTGAGGCAAAGAC
ATTGGGAAAATGGGTGCGGAGATCCGTTTACCCAGGAACCGAACCCGACTTTGGCTCG
GAACATTCGACACGGCTGAAGAAGCCGCGTTGGCTTATGATCGCGCCGCGTTTAAGCTTC
GTGGTGACTCGGCTCGGCTTAACTTCCCAGCTCTCCGATACCAAACCGGCTCGTCTCCGT
CTGATACCGGCGAATATGGTCCTATTCAAGCTGCCGTAGACGCTAAACTAGAAGCCATAT
TAGCTGAGCCGAAGAATCAGCCGGGCAAAACGGAGAGGACGTCGAGGAAACGAGCTAAAG
CCGCGGCTTCTTTCAGCTGAGCAGCCGTCAGCGCCACAACAACATTCCGGGTGCGGTGAAA
GTGATGGGTGCGGTTTACCGACTTCCGATGTTATGGTGCAGGAGATGTGCCAAGAGCCAG
AGATGCCATGGAATGAAAATTTATGCTCGGCAAGTGTCTTCTTATGAGATAGATTGGG
CTTCAATTTTATCGTGAAAAATTAGGATTCAATTCATTTTTATTTCATTTTAACTTGTTTG
TATTTTCTTTTAACTTTAGGGTTATTAGCTGTGCGTAA
>G1384 Amino Acid Sequence (domain in AA coordinates: TBD)
MADLFGGGHGGELMEALQPFYKSASTSASNPAFASSNDAFASAPNDLFSSSSYYNPHASL
FPSHSTTSYDPIYSGSMTYPSSFGSDLQQPENYQSQPHYQNTITYTHQDNNTCMLNFIEP
SQPFMTQPGPSSGSVSKPAKLYRGVVRQRHWGKWA EIRLPRNRLRLWLGTDFDTAEAAAL
AYDRAAFKLRGDSARLNFALRYQTGSSPSDTGEYGP IQAAVD AKLEAILAEPKNQPGKT
ERTSRKRAKAAASSAEQPSAPQQHSGSGESDGS GSPTS DVMVQEMCQEP EMPWNENFMLG
KCPSYEIDWASILS*

>G1399 (261..1475)

AGGTCGAATTTTCTGAAATTAAGATTCAATTCCTCCATGGAAGAAGCTCTGTTTTTATTCT
CTTTAGCTTAGCTTAGCTTCTACTGATCTGTTTTTGCTACAAAATCCCATCTTTTTCTTT
AAAACCTTTTATCTCTGAATCTTGAGTTTCTTGTAAGAAGAAGCAATTTGAATCTTT
CGTAATCATAAAGATTCTGGAGGATCTCTACTGATTTGTGCGAATCTCTCACTACAGAA
TCACTTGATCTTATGTCCGGATGGAGGAGAGAGAAGGAACCAACATCAACAACAACATCA
CTAGCAGTTTCGGCTTGAAGCAGCAACATGAAGCTGCTGCTTCTGATGGTGGTTACTCAA
TGGACCCACCACCAAGACCCGAAAACCTAACCCTTTTGTAGTCCACCCCACTACTGTCC
CCGCGGCCGCCACCGTAGCAGCAGCTGTTACTGAGAATGCGGCTACTCCGTTTAGCTTAA
CAATGCCGACGGAGAACACTTCAGCTGAGCAGCTGAAAAAGAAGAGAGGTAGGCCGAGAA
AGTATAATCCCGATGGGACTTGTGTCGTGACTTTATCGCCGATGCCAATCTCGTCTCTG
TTCCGTTGACGTCGGAGTTTCTCCAAGGAAACGAGGAAGAGGACGTGGCAAGTCTAATC
GATGGCTCAAGAAGTCTCAAATGTTCCAATTCGATAGAAGTCTGTTGATACCAATTTGG
CAGGTGTAGGAAGTCTGATTTTGTGTTGGTGCCAACTTTACACCTCATGTACTGATCGTCA
ACGCCGGAGAGGATGTGACGATGAAGATAATGACATTCTCTCAACAAGGATCTCGTGCTA
TCTGCATCTTTTCACTAATGGTCCCATCTCCAATGTTACGCTTCGTCAATCTATGACAT
CCGGTGGTACTCTAATTTATGAGGGTCTTTTTGAGATTCTCTCTTTGACGGGTTCTGTTA
TGCAAAATGACTCTGGAGGAACCTCGAAGTAGAGCTGGTGGTATGAGTGTTCCTTGCAG
GACCAGATGGTCTGTCTTTGGTGGAGGACTCGCTGGTCTCTTTCTTGCTGCTGGTCCG
TCCAGGTAATGGTAGGGACTTTTATAGCTGGTCAAGAGCAGTCACAGCTGGAGCTAGCAA
AAGAAAGACGGCTAAGATTGGGGCTCAACCATCTTCTATCTCTTTAACATATCCGCAG
AAGAACGGAAGGCGAGATTCGAGAGGCTTAACAAGTCTGTTGCTATTCTGCACCAACCA
CTTCATACACGCATGTAAACACAACAATGCGGTTTACAGTTACTATACAACTCGGTTA
ACCATGTCAAGGATCCCTTCTCGTCTATCCAGTAGGAGGAGGAGGAGGTGGAGAGGTAG
GAGAAGAAGAGGGTGAAGAAGATGATGATGAATTAGAAGGTGAAGACGAAGAATTCGGAG
GCGATAGCCAATCTGACAACGAGATTCCGAGCTGATGATGATCATACGGTTTCTTTTCGC
GGATTTGTTAGGTTTGTGATGATTTTCAAGATTTTGGTTGATTGTTTTTATTAACACAGAATG
TTTAGAAGCTGCTATCTTTAGGTTCCCATCTCTTGTGATTGTTGAGTATCCTTGTGTA
AACAACTTACTGTTGCAAACTCTCTTCAAAAAAGTTTCACTTTGCTTTCCCA

>G1399 Amino Acid Sequence (domain in AA coordinates: 86-93)

MEEREGTNINNNITSSFGLKQHEAAASDGGYSMDPPRPENPNPFLVPPTTVPAATVA
AAVTENAAATPFSLTMPTEENTSABQLKKRGRPRKYNPDPGLVVTLSPMPISSSVPLTSEF
PPRKRGRGRGKSNRWLKKSQMFQFDRSPVDTNLAGVGTADFGVANFPHVLIVNAGEDVT
MKIMTFSSQGSRAICILSANGPISNVTLRQSMSTSGGTLTYEGRFEILSLTGSFMQNDSSG
TRSRAGGMSVCLAGPDGRVFGGLAGLFLAAGPVQVMVGTFIAGQEQSLELAKERRLR
GAQPSSISFNISAEERKARFERLNKSVAIAPPTTSYTHVNTTNAVHSYYTNSVNVHKDPF
SSIIPVGGGGGGEVGEEDDELEGEDEFGGDSQSDNEIPS*

>G1415 (60..680)

CTTTATCACTCACCAAAAGTCGTACATAATATCACTTTTCGAGTTATCAACATCCGTACA
TGTCATCCATAGAGCCAAAAGTAATGATGGTTGGTGCTAATAAGAAACAACGAACCGTCC
AAGCTAGTTTCGAGGAAAGGTTGTATGAGAGGAAAAGGTGGACCCGATAACGCGTCTTGCA
CTTACAAAGGTGTTAGACAACGCACTTGGGGCAAATGGGTCGCTGAGATCCGCGAGCCTA
ACCGAGGAGCTCGTCTTTGGCTCGGTACCTTCGACACCTCCCGTGAAGCTGCCTTGGCTT
ATGACTCCGCGAGCTCGTAAGCTCTATGGGCCTGAGGCTCATCTCAACCTCCCTGAGTCCT
TAAGAAGTTACCCTAAACGCGCTCGTCTCCGCGCTCCAGACTACACCAAGCAGCAACA
CCGGTGGAAAAAGCAGCAGCGACTCTGAGTCGCCGTGTTTCATCCAACGAGATGTCATCAT
GTGGAAGAGTGACAGAGGAGATATCATGGGAGCATATAAACGTGGATTGCGCGTAATGG
ATGATTTCTCAATATGGGAAGAAGCTACAATGTCGTTAGGATTTCCATGGGTTTCATGAAG
GAGATAATGATATTTCTCGGTTTGATACTTGTATTTCCGGTGGCTATTCTAATTGGGATT
CTTTTCATTCCCCACTTTGAGGTGTCACTAGACTCTCTTTAATTGTTAAGTTATCATATA
CAAACATACATATATATACAAATATAGTCACCGTGAAGTATAGTAAATAAACA
CCAGTTACATGTACTTATATGTGCACATCTATATATGTGTTTGTCTGTATAGTGTGA
AAGCAGATTCTTACCATATCA

>G1415 Amino Acid Sequence (domain in AA coordinates: TBD)

MSSIEPKVMVGANKKQRTVQASSRKGCMRGKGGPDNASCTYKGVQRQTWGKWVAEIREP
NRGARLWLGTFTDSREAAALAYDSAARKLYGPEAHLNLPESLRSPKTPASSPASQTPSSN
TGGKSSSDSESPSSNEMSSCGRVTEEISWEHINVDLPVMDSSIWEEATMSLGFPWVHE

GDNDISRFDTCISGGYSNWDSPHSP*

>G1417 (32..1501)

TCTATCTCTATCTATCTCTCTTTGTCTGCAATGGAAGAACATATTCAGATCGCCGTGA
AATTGCGTTCTTACACTCAGGAGAATTTCTCCACGGAGATTCTGACTCAAAGGATCATCA
ACCGAACGAGTCTCCGGTGAACGTCATCAGAGTCGTCTATCAAAGAAGTTGATTTCTT
CGCTGCTAAAAGTCAGCCGTTTGATCTTGGTCATGTGAGAACAACGACGATCGTTGGATC
ATCTGGTTTTAATGATGGATTAGGTTTGGTAAATTCATGTCATGGAACATCAAGCAATGA
TGGCGATGACAAAACCAAACCTCAAATTAGTAGACTGAAGTTGGAGCTAGAGAGGCTTCA
CGAGGAGAATCACAACCTGAAGCATTATTAGATGAGGTGAGTGAGAGTTACAACGACCT
CCAAAGAAGAGTTTTGTAGCAAGACAAACACAAGTGAAGGTCTTCATCATAAACACA
TGAGGATGTACCTCAAGCTGGTTCCTCACAAGCTCTAGAGAACAGAAGACCAAAGGATAT
GAACCATGAAACTCCGGCCACCACCTTGAAACGACGGTCTCCAGACGACGTGGATGGTCG
TGATATGCACCGAGGATCACCAAAAACCTCCTCGAATAGACCAAAACAAGAGTACTAATCA
TGAAGAACAACAAAACCTCATGATCAATTACCTTATAGAAAAGCTAGGGTTTCCGTTAG
AGCTAGATCTGATGCCACTACGGTAAATGACGGATGTCAATGGAGAAAATACGGTCAGAA
AATGGCGAAAGGGAATCCATGTCTCGCGCTTATTATCGTTGCACCATGGCCGTTGGATG
TCCTGTCCGTAAACAGGTCCAACGATGCGCGGAGGATACAACCTATCTTGACAACAACGTA
CGAAGGAAACCATACCATCTCTTCCCCCGTCAGCCACAGCCATGGCTGCAACCACCTC
CGCCGAGCAGCCATGCTCTTATCAGGCTCCTCTCCAGCAACCTCCACCAAACACTCTC
TAGCCCCCTCCGCCACGTATCATCATCTTCTACCATAACTTCCCATACACCTCCACAAT
CGCAACACTCTCTCGCTCAGCTCCTTTCCCCACCATAACCTTAGACCTCACCAACCCACC
TCGACCGCTACAACCGCCACCGCAGTTTCTAAGCCAGTATGGTCCC GCCGCTTTTTTACC
AAACGCTAATCAAATTAGGTCTATGAATAATAATAACCAGCAGTTATTAATACCTAATTT
GTTTGGCCCAAGCCCCACCACGTGAAATGGTCGATTAGTTAGGGCTGCGATTGCGAT
GGATCCGAACCTTACGGCGGCACCTTGCGGCCGCGATCTCAAACATTATCGGAGGAGGTAA
TAACGACAACAATAATAACTGATATTAATGATAACAAGGTTGATGCAAAAAGTGGAGG
GAGTAGTAACGGAGATTGCCACAGCTTCTCAGTCTTGCACCACCTTCTCTACAAACTA
ATTTTACTACCATTATTATATGTTATCTTATTATATATATACACACATATTATACATTA
TGCGTATCTTAAGTTTTTTTTTTGGGGCCATTATATATGAATGATATGGAGATCACTGAG
AGAGAGAGAGAGCTATTATGGGTTTTTTTTT

>G1417 Amino Acid Sequence (domain in AA coordinates: 239-296)

MEEHIQDRREIAFLHSGEFLHGDSKDHQPNESPVERHHESSIKEVDFFAAKSQPFDLG
HVRTTTIVGSSGFNDGLGLVNSCHGTSSNDGDDKTKTQISRLKLELERLHEENHKLKHL
DEVSESYNDLQRRVLLARQTQVEGLHHKHQEDVPQAGSSQALENRRPKDMNHETPATTLK
RRSPDDVDGRDMHRGSPKTPRIDQNKSTNHEEQNPQDQLPYRKARVSVRARSDDATTVND
GCQWRKYGQKMAKGNPCPRAYRCTMAVGC PVRKQVQRCAEDTTILTITYEGNNHPLPP
SATAMAAATTSAAAAMLLSGSSSNLHQTLSSPSATSSSSFYHNFYPTSTIATLSASAPFP
TITLDLTNPRLPQPPQFLSQYGPAFLPNANQIRSMNNNNQQLLIPNLFGPQAPPREM
VDSVRAAIAMPNFTAALAAISNIIGGNNDNNNTDINDNKVDAKSGSSNGDSPQLP
QSCTTFSTN*

>G1442 (1..1293)

ATGGGAACAAGAGCAGAACGCAAGGAAGATTTTGTGGTGGGTTTGGATTGGTGTGTA
GAAAATTCGCATAAAGACGTTATGGTGCTACCTCATCATCACTATTATCCATCATATTCA
TCACCTTCCTCTTCTTTGTGTTACTGTTCTGCTGGTGTAGCGATCCCATGTTCTCT
GTTTCTAGCAATCAGGCTTACACTTCTTCTCACAGTGGTATGTTACACCCGCCGTTCT
GGTCTGCTGCTGTGACTGTAGCAGATCTTTTTTCTCCTTGAGCTCTCAGGGGAAATG
AGAAGAAGTATGAACGAAGATGCTGGTGCAGCTTTCAGCGAAGCTCAATGGCATGAGCTT
GAGAGGCAGAGGAATATATACAAGTACATGATGGCTTCTGTTCTCTGTTCTCCAGAGCTT
CTCACACCTTTCCCAAGAACCACCAATCAAACACTAACC CGGATGTAACCTGTGGCAGTG
GCGACAGGAGGCTCAFTGCAGCTGGGGATTGCTTCAAGCGCAAGCAATAACACGGCTGAT
CTGGAGCCATGGAGGTGCAAGAGAACAGATGGGAAGAAATGGAGGTGCTCTAGAAACGTG
ATTCTGATCAGAAATACTGTGAGAGACACACACAAGAGCCGCTCTCGTTCAAGAAAG
CATGTGGAATCATCTACCAATCATCTCACCACAATGACATTCTGACGCTAAGAATGAT
ACTAGCCAGCTTGTGAGAACTTATCCTCAGTTTTACGGACAACCTATAAGCCAGATCCCT
GTGCTTTCTACTCTTCCGTCTGCCTCCTCTCCATATGATCACCACAGAGGACTGAGGTGG
TTTACGAAAGAAGATGATGCCATTGGAACCTTAAACCCGGAGACTCAAGAAGCTGTCCAG
CTGAAAGTTGGATCAAGCAGAGAGCTCAAACGGGGATTGATTATGATCTGAATTTACAG

CAGAAAGAGCCAATAGTAGACCAGAGCTTTGGAGCATTGCAGGGTCTATTAAGTCTAAAC
CAGACACCACAACATAACCAAGAAACAAGACAGTTTGTGTAGAAGGAAAGCAAGATGAA
GCGATGGGAAGCTCTCTGACACTCTCAATGGCTGGAGGAGGCATGGAGGAAACAGAGGGA
ACAAACCAGCATCAGTGGGTTAGCCATGAAGGTCCATCATGGCTCTATTCAACAACACCA
GGTGGACCATTTGGCTGAAGCACTGTGTCTCGGTGTCTCCAACAACCCAAGTTCTAGTACT
ACTACTAGTAGCTGCAGCAGAAGCTCAAGCTAA

>G1442 Amino Acid Sequence (domain in AA coordinates: 172-223)
MGTRAERKEDFVGFGFGVVENSHKDVMLPHHHYPSYSSPSSSSSLCYCSAGVSDPMFS
VSSNQAYTSSHSGMFTPAGSGSAAVTVADPFFSLSSSGEMRRSMNEDAGAAFSEAQWHEL
ERQRNIYKYMMASVPVPPELLTPFPKNHQSNTNPDVTVAATGGSLLGLIASSASNNTAD
LEPWRCRKTGKKWRCNRNVIDQKYCERHTRKSRPRSRKHVESSHQSSHHNDIRTAKN
TSQLVRYTYPQFYGPISQIPVLSLTPSASSPYDHRGLRWFTKEDDAIGTLNPETQEA
VQ LKVGSSRELKRGFDYDLNFRQKEPIVDQSPGALQGLLSLNQTPQHNOETROFVVEGKQDE
AMGSSSLTLMSAGGMEETEGTNQHQVWSHEGPSWLYSTTPGGPLAALCLGVSNNPSSST
TTSSCSRSSS*

>G1454 (86..1180)
CTAGTAGTGATGATATGATCGCTTCTTCTCCTACAATCTCAGAAACCTCCGATCACGGTT
TTAGATATCTTCTACAACGGATACAATGGAGAGCACCGATTCTTCCGGTGGTCCACCACC
GCCACAACCTAACCTTCTCCAGGCTTCCGGTTTACCCTACCGACGAAGAGCTTGTGT
TCACTACCTCAAACGCAAAGCAGCCTCTGCTCCTTTACCTGTGCGCCATCATCGCCGAAGT
CGATCTCTATAAATTGATCCATGGGAACCTCCCGCTAAAGCATCGTTTGGAGAACAAGA
ATGGTACTTCTTTTACGAGATCGGAAGTATCCAAACGGAGCAAGACCAACAGAGC
GGCGACTTCAGGTTATTGGAAGCGACCGGTACAGATAAACCGGTACTTGCTTCCGACGG
TAACCAAAAGGTGGGCGTGAAGAAGGCACTAGTCTTCTACAGTGGTAAACCAACAAAGG
CGTTAAAGTGATTGGATCATGCATGAGTATCGTCTCATCGAAAAACAAACAAATCG
ACCTCTGGCTGTGATTTCGGCAACAAAAAACTCACTCAGACTTGATGATTGGGTGT
ATGTAGAATCTACAAGAAGAACACGCAAGTCGACATGTTGATAACGATAAGGATCATGA
TATGATCGATTACATTTTCAAGAAAGATTCTCCGCTTTTATCAATGGCGGCTGCTTCTAC
AGGACTTCACCAACATCATATAATGTCTCAAGATCAATGAATTTCTTCCCTGGCAAATT
CTCCGGTGGTGGTTACGGGATTTTCTCTGACGGTGGTAACACGAGTATATACGACGGCGG
TGGCATGATCAACAATATTGGTACTGACTCAGTAGATCACGACAATAACGCTGACGTCGT
TGGTTTAAATCATGCTTCGTCGTGAGTCCATGATGATGGCGAATTGAAACGAACCTCT
CCCAGTGGCGTATTGGCCTGTAGCAGATGAGGAGCAAGATGCATCTCCGAGCAAACGGTT
TCACGGTGTAGGAGGAGGAGGAGGAGATTGTTCGAACATGTCTTCTCCATGATGGAAGA
GACTCCACCATTGATGCAACAACAAGGTGGTGTGTTAGGAGATGGATTATTCAGAACGAC
ATCGTACCAATTACCCGGTTAAATTGGTACTCTTCTTAATCAAATGTGTTTCGCCGCCG
GTGTGAAGAATTTTCCGGTGACAGTGAAGATTTTTTTCCGATTGGTGGGGTCATTTCAT
GCATTATATAATTTGAGATTTGTGTATATGTTTTGGGTTAATTAATTGGTCACAGGGG
>G1454 Amino Acid Sequence (conserved domain in AA coordinates: 9-178)

MESTDSSGGPPPPQPNLPPGFRFHPTDEELVVHYLKRKAASAPLPVAIAEVDLYKFDPW
ELPAKASFGGEQWYFFSPDRKYPNGARPNRAATSGYWKATGTDKPVLASDGNQKVGVKK
ALVFYSGKPKGVKSDWIMHEYRLIENKPNRPPGCDGFKNKNLSRLDDWVLCRIYKKN
ASRHDNDKDHMDIYIFRKIPPSLSMAAASGLHQHHNVSRSMNFFPGKFSGGGYGIF
SDGGNTSIYDGGMINNIGTDSVDHDNNADVGLNHASSSGPMMANLKRTPVPYWPVA
DDEQDASPSKRFHVGVGGGGDCSNMSSSMMEETPPLMQQGGVLDGLFRFTTSYQLPGLN
WYSS*

>G1459 (1..1272)
ATGATGAAAGGTCTGATTGGGTATAGATTTAGTCCGACGGGAGAGGAAGTGATCAACCAT
TACCTAAAGAACAACTTCTGGGTAAAGTATTGGCTCGTTGATGAAGCTATTAGCGAGATC
AACATCTTGAGTCACAAACCCAGCAAGGATTTGCCTAAGTTAGCTAGGATCCAATCGGAA
GATCTTGAATGGTATTTCTTCTCCGATTGAGTACACGAACCCGAATAAGATGAAAATG
AAGAGGACGACAGGTTCTGGGTTTTGGAACCTACTGGTGTGATCGGGAAATTAGGGAT
AAAAGAGGAAATGGTGTGTGATAGGATTAAGAAGACGCTTGTGTACCATGAAGGTAAG
AGTCCTCATGGAGTTAGAATCCTTGGGTTATGCACGAGTATCATCACTTGCTTGCTT
CATCATAAGAGGAAATATGTTGTCTGCCAAGTAAAGTATAAGGGTGAAGCTGCAGAAATT
TCATATGAGCCAAGTCCCTCTTTGGTATCCGATTGCATACCGTCATAGCGATTACCGGA
GAACCGGAACCTGAGCTTCAGGTTGAGCAGCCAGGTAAAGAAAATCTCTTGGGTATGTCT

GTAGATGATTTGATAGAACCAATGAACCAACAAGAGGAGCCACAAGGTCCTCACCTTAGCT
CCGAATGATGATGAGTTTATACGTGGATTGAGGCATGTTGATCGAGGGACGGTTGAATAT
TTGTTTGCCAATGAAGAAAACATGGATGGTTTGTCTATGAATGACTTGAGAATCCCAATG
ATCGTCCAACAAGAGGATCTCTCTGAGTGGGAGGGATTTAACGCAGACACCTTTTTTCAGC
GACAACAACAATAACTATAACCTTAACGTGCATCATCAACTAACGCCTTACGGCGATGGC
TATTTGAATGCATTTTCGGGTTATAACGAAGGGAATCCTCCCGATCACGAATTAGTGATG
CAAGAGAACCGCAACGATCACATGCCAAGGAAACCTGTGACAGGGACCATTGATTATAGC
AGCGATAGTGGCAGTGATGCTGGATCCATATCTACAACGGTGAAACAAGAAATCCCAAGA
GCTGTTGATGCACCCATGAACAATGAGTCATCTTTGGTGAAAACAGAGAAGAAAGGCTTG
TTTATTGTAGAGGACGCAATGGAGAGAAACCGCAAGAAACCACGATTATCTATCTCATG
AAGATGATCATAGGCAACATCATATCGGTTTTACTACCCGTCAAAGATTGATCCCGGTG
AAGAAGTTATGA

>G1459 Amino Acid Sequence (conserved domain in AA coordinates:10-152)

MMKGLIGYRFSPTGEEVINHYLKNKLLGKYWLVDIAISEINILSHKPSKDLPLKARIQSE
DLEWYFFSPIEYTNPNKMKMKRTTSGGFWKPTGVDREIRDKRGNGVVIGIKKTLVYHEGK
SPHGVRTFPWMHEYHITCLPHHKRKYVVCQVKYGEAAEISYEPSPSLVSDSHTVIAITG
EPEPELQVEQPGKENLLGMSVDDLI EPMNQEEPPQGPHLAPNDDEFIRGLRHVDRGTVEY
LFANEENMDGLSMNDLRIPIVQQEDLSEWEGFNADTFSDNNNNYNLNVHQLTPYGDG
YLNAFSGYNEGNPPDHELVMQENRNDHMPKPVGTIDYSSDSGSDAGSISTTVKQEI PR
AVDAPMNNESLVLKTEKGLFIVEDAMERNRKKPRFIYLMKMIIGNIISVLLPVKRLIPV
KKL*

>G1460 (87..995)

CGTCGACCTTCACTCAAACCTAATCCCGGAACCCGGGAATTTTGATCATTTTGTCTTCT
TTTCGATCTGTTTCTATTTTAAAAAGATGATGAAAGATCCGACTGGGTATAGATTTAGTC
CGACGGGAGAGGAAGTGATAAACCATTAACCTAAAGAAACAAATTCGGGTAAGACTTGGC
TCGTTGATGAAGCCATTAGCGAGATCAACATCTTGAATCACAACCCAGCAAGGATTTGC
CTAAGTTAGCTAGGATCCAATCGGAAGATCTTGAGTGGTACTTTTTCTCTCCGATTGAGT
ACACGAACCCGAATAAGATGAAAAATGAAGAGGACGACAGGTTCTGGGTTTTGGAACCTA
GTGGTGTGATCGGAAAATTAGGGATAAAAGAGGAAATGGTGTGTGATAGGGATTAAGA
AGACGCTTGTGTACCATGAAGGTAAGAGTCTTCATGGAGTTAGAACTCCTTGGGTTATGC
ACGAGTATCACATCACTTGCTTGCCCTCATATAAGAGGAAATATGTTGTCTGCCAAGTAA
AGTATAAGGGTGAAGCTGCAGAAATTTTATATGAGCCAAGTCCCTCTTGGTATCCGATT
CGCATACCGTCATAGCGATTAACGGAGAACCGGAACCTGAGCTTCAGGTTGAGCAGCCAG
GTAAAGAAAATCTCTTGGGTATGTCTGTAGATGATTGATAGAACCAATGAACCAACAAG
AGGAGCCACAAGGTCTCACTTAGCTCCGAATGATGATGAGTTTATACGTGGATTGAGAC
ATGTTGATCGAGAGCCGGTTGAATATTTGTTTGCCAATGAAGAAAACATGGATGGTTTGT
CTATTATGAATGACTTGACAAATCCCAATGATCGCCCAACAAGAGGATCTCATTCTCTCTG
AGTGGGAGGGATTTATCGCAGCCACCTTTTTTCAGCGACAACAACAATAACAATAACCTTA
ACGTGCATCAACTAACGTCTTTCTTACCGGGATGATTATCAGAATGCATTTTGGGTTACA
ACGGAGCGNCCGCT

>G1460 Amino Acid Sequence (domain in AA coordinates: TBD)

MMKDPTGYRFSPTGEEVINHYLKNKILGKTWLVDEAISEINILNHKPSKDLPLKARIQSE
DLEWYFFSPIEYTNPNKMKMKRTTSGGFWKPSGVDRKIRDKRGNGVVIGIKKTLVYHEGK
SPHGVRTFPWMHEYHITCLPHHKRKYVVCQVKYGEAAEISYEPSPSLVSDSHTVIAING
EPEPELQVEQPGKENLLGMSVDDLI EPMNQEEPPQGPHLAPNDDEFIRGLRHVDREPVEY
LFANEENMDGLSIMNDLTIPIAQQEDLILSEWEGFIAATFFSDNNNNNNLNVHQLTSFL
PG*

>G147 (37..672)

AAATCATCAGATAGAAGGAAATATTCTGATTGAGAGATGGCTCGTGGAAAGATTGAGCTT
AAGAGGATTGAGAACCCTGTTACAGACAAGTGACTTTTTCAGAGGAGAACTGGTCTT
CTCAAGAAGGCTAAGGAGCTCTCTGTGCTGATGCCGAGATCGGTGTTGTGATCTTC
TCTCCTCAGGGCAAGCTCTTTGAGCTCGCTACTAAAGGAACAATGGAGGGAATGATTGAT
AAGTACATGAAGTGACTGGTGGTGGTGGTGGTCTGCTTCTTCTGCTACTTTTACTGCTCAA
GAACAACCTTCAACCACCAATCTTGATCCGAAAGATGAGATCAACGTGCTTAAGCAAGAG
ATTGAGATGCTTCAGAAAGGGATAAGCTATATGTTTGGAGGAGGAGATGGGGCTATGAAT
CTTGAAGAACTCTTTTGGCTTGAGAAGCATCTTGAGTATTGGATTTCTCAGATTGCTCT
GCTAAGATGGATGTTATGCTTCAAGAAATTCAGTCATTGAGGAACAAGGAAGGAGTCCTC

AAAAACACCAACAAGTATCTCCTCGACAAGATAGAGGAAAAACAACATAGCATATTAGAT
GCTAACTTCGCAGTCATGGAGACAACTATTCCTATCCGCTAACAATGCCAAGTGAAATA
TTTCAGTTCCTAGACCATAGGGTATTTGAAGACTATGTCTCACGAATTTAAATAACCTTGG
TAAGTATAATATAGTGTGTAAATCACACATAATTAAAAATAAGCCTGTGGAACCTCGC
TAGGCAGTTGAAAAATCTATCCGTATGTTTATCCTCTTGTTTACATTTGTTGGTGTGAA
GATGAAATGACTGCAAGTGTGGTGTGTACTTATAACTCTTCTACTTTCTATCTATGTTT
TGAATTTATGGATT

>G147 Amino Acid Sequence (domain in AA coordinates: 2-57)
MARGKIQLKRIENPVHRQVTFCKRRTGLLKKAKELSVLCDAEIGVVIFSPQGLFELATK
GTMEGMIDKYMKCTGGGRGSSSATFTAQEQQLQPPNLDPKDEINVVKQEIEMLQKGISYMF
GGGDGAMNLEELLLLEKHLEYWISQIRSAKMDVMLQEIQSLRNKEGVLKNTNKYLLDKIE
ENNNSILDANFAVMEITNYSYPLTMPSEIFQF*

>G1471 (1..735)
ATGGAGAACCAATCTATGTCTTCATCAAGCTCCTCCACACACAAACATGATCAAAAACTC
AAAAGTTCCGTTGTGGCCATGGAGGTCTGGAGGAGAAGGAGACAGTGAACAATCCGCCC
CAGTATTATAATAAGATCTACATCTGTTACTTGTGCAAGAGAGCGTTCCCAACCCCTCAT
GCCCTTGGCGGTACGGAACCAACCCACAAGGAGGACCGAGAATTGGAGAGGCAACAGATC
GAGTCAAGGCTTTCTAACAAAGACAAGTCTAACTTGCTCTTTGGTGGGTCTTCACAAGAT
GTTTATCAAATGATAATCACCTTGGACTCTCTCTTGGTCCATTGAAGTCCATAGAAGGT
AGCAGCAGCAGCAACAACGTTAACCATTGCTTAATGTTGGAGTCCCTAGAGGAACCA
GATATGAACATGAACAACATATAGCTCACATGCTTTATCAACTGATGATATTAATCTTGAT
CTTACTCTTGGTCCATCTAAGTCCATAGGAGATAGCAACAATATCATTAATAACAACACT
AACTCATCCTTCGATGGGAATCTGATCATTCCTCGTTCGTCTCGTGTGTCTAGATACCAT
TTTGTGCTGGGAACCCCTTGATTCAATCTCTAGAAACATTCTCTCTTCTATTACTTTT
CCTCATCTAAACATCAATCTTCTCATGATTCGTTTCTTTACAAGAGAATGGTTCGGGC
TCTAGTCACTCATAA

>G1471 Amino Acid Sequence (domain in AA coordinates: 49-70)
MENQSMSSSSSTHKHDQKLKSSVVAMEVLEEKETVMNPPQYINKIYICYLCKRAFPTPH
ALGGHGTTHKEDRELERQQIESRLSNKDKSNLLFGGSSQDVLSDNHLGLSLGPLKSIEG
SSSSNNVNPPLLNVGVPRGTTDMNMNYSHALSTDDINLDLTLGPSKSIKGSNNIINNNT
NSSFDGNLIIPVRPRVSRVYHFVAGNPLDSISRNIPPSITPHLNINLSHDSFSLQENGSG
SSHS*

>G1475 (1..645)
ATGAAGAGAACACATTTGGCAAGTTTTAGTAACAGAGACAAAACCCAGAAGAAGAAGGA
GAAGACGGTAATGGTGACAACAGAGTCATCATGAATCACTACAAGAATTACGAAGCTGGG
CTGATCCCATGGCTGCCAAGAATTACACTTGCAGCTTCTGCAGGAGAGAGTTTCAATCT
GCTCAAGCACTTGGAGGCCACATGAATGTTTATAGAAGAGACAGAGCAAAACTCAGGCAG
ATCCCTTCTTGGCTCTTTCGAACCTCACCAACACACACCTATTGCAAACCCCTAACCTAAT
TTTAGCTCTTCTTCTCTCTTCAACAACAACAGCTCATCTTGAGCCTTCCCTAACCAAC
CAGAGATCCAAAACCACTCCTTTCTCTTCTGCCCCGTTTGATCTTTTGGACAGTACTACT
AGCTATGGAGGTTTGATGATGGACAGAGAGAAGAACAAGAGCAATGTATGTAGCAGAGAG
ATCAAGAAAAGTGCCATCGATGCATGTCATTAGTAAGATGTGAGATAAGCCGTGGGGAT
CTGATGAATAAGAAAGATGATCAAGTCATGGGTTGGAGCTTGGGATGAGTTTGAGGAAT
CCCAACCAAGTTCTTGATTGGAGCTTCGACTAGGCTACCTCTAA

>G1475 Amino Acid Sequence (domain in AA coordinates: 51-73)
MKRTHLASFSNRDKTQEEEGEDGNDNRVIMNHYKNYEAGLIWPPPKNYTCSFCRREFRS
AQALGGHNMVHRRDRAKLRQIPSWLFEPHHHTPIANPNPNFSSSSSSSTTAHLEPSLTN
QRSKTTFFPSARFDLDDSTTSYGLMMDREKNKSNVCSREIKKSAIDACHSVRCEISRGD
LMNKKDDQVMGLELGLMSLRNPNQVLDLELRLGYL*

>G1477 (1..606)
ATGTTGTCTCGGACTCGAATTACGCTAGTGATATTAGCGACGATGCCCTCCGCCACCGGA
TCGATAGAGAATCCTATATACAAATGCAAGTATGTCTAGGAAGTTGATAAAACACAA
GCATTAGTGGTCATCAAAATGCACAGAAAGGAGAGAGAGGTGAAAAACACAAAAA
GCATTTTGGCGCATTTGAACCGACCAGAACCAGATCTTACGCGTACTCGTATTCTGAT
CATCATTCATTTCTAACCAATACGCACTCCACCGGGATTGTAACAGCCTCAGTACAAA
GTTGATAGATCATACAAGATGTCCATGGTCTACAACCAATATGTGGGATCCTCAAGCTCT
AGCTTTGCAGGACTACAAAGTGACCAAGTCAAGGAATGAACCAGGATTGGACCTTTACC

GGGATCCCATTTCCTACCCCAATCTCAACCTCAACCACTATCGTCACCAATATGTTTGGAT
CTTTGCCTTGGCATTGGTAGCTCCCAAACCAACCACAACCTCAAGAACCAATGATGCA
ACAGAAGAGATGGATGCTGAGAAAGAAAATGATGGTTCTTCCCTTTCTCTCTCACTCAAA
CTGTGA

>G1477 Amino Acid Sequence (domain in AA coordinates: 29-48)
MLSSDSNYASDISDDASATGSIENPIYCKYCPKFKDTQALGGHQNAHRKEREVEKQKQK
AFLAHLNRPEPDLYAYSYSYHHSFPNQYALPPGFEPQYKVDRSYKMSMVYNQYVGSSSS
SFAGLQSDPSQGMNQDWTFTGIPFLPQSQPQLSSPICLDLCLGIGSSQTQPPQPEPND
TEEMDAEKENDGSSLSLSLKL*

>G1487 (1..1020)
ATGGAACAAGCCGCTTGAAGAGCAGCGTCAGGAAAGAGATGGCTCTCAAAACGACTTCT
CCGCTTTACGAAGAGTTTCTTGCCTGACCACCGCTCAAAATGGCTTTTCCGTCGACGAT
TTCTCTGTAGACGACTTGCTTGACTTGTCAAACGATGACGTTTTTGGCGACGAAGAACT
GACCTCAAGGCTCAACATGAGATGGTCCGTGTTTCTCTGAGGAACCAACGACGACGGA
GACGCTCTTCCGCGGAGCAGCGATTTCTCCGGCTGTGACGACTTTGGTTCTCTCCCTACA
AGCGAACTCTCTCTTCCGGCGGATGATTTAGCGAACCTTGAGTGGCTCTCTCATTTCGTG
GAGGACTCCTTACGGAATATTCGGGTCCAACCTCACCAGAACCCGACTGAGAAACCG
GCGTGGTTAACGGGTGACCGGAAACATCTGTGACTGCAGTCACGGAAGAGACCTGTTTC
AAATCCCTGTTCCGGCTAAAGCCCGTAGCAACGTAACCGCAATGGCCTCAAGGTCTGG
TCGCTTGGTTTCGTCTGCTCCTCCTCGGGTCTTCTCTGTCGGTTTCGACCTCCTCCTCTCT
TCGGGTCTTCCAGCCCGTGGTTCTCCGGCGCTGAGCTGCTCGAGCCTGTGGTCACGTCA
GAGAGGCCACCGTTTCCCAAGAAGCATAAGAAAAGGTGAGCCGAGTCTGTTTTCTCCGGT
GAGCTGCAGCAGCTGCAACCTCAGCGAAAGTGCAGCCACTGCGGCGTTTCAGAAAACCTCCG
CAGTGGAGAGCCGGGCCAATGGGAGCCAAGACCCTGTGCAATGCGTGCAGGTGTCCGGTAC
AAGTCGGGTAGGTTGCTACCGGAATACAGACCCGCTTGTAGCCCGACATTCTCGAGTGAG
CTGCACTCGAACCACCACCGGAAAGTCATAGAGATGAGGCGGAAGAAGGAGCCAACCACT
GACAACGAAACCGGTTTAAACCAGCTGGTTCACTCCCCACAAGCTGTACCAAGTTTTTGA

>G1487 Amino Acid Sequence (domain in AA coordinates: 251-276).
MEQAALKSSVRKEMALKTTSPVYEEFLAVTTAQNFGSVDDFVDDLLDLNDDVFADERT
DLKAQHEMVRVSSEEPNDGDALRRSSDFSGCDDFGSLPTSELSLPADDLANLEWLSHFV
EDSFTEYSGPNLTGTPTEKPAWLTGDRKHPVTAVTEETCFKSPVPAKARSKRNRNGLKVW
SLGSSSSSGPSSSGSTSSSSSGPSSPWFSGAELEPVVTSERPPFPKHKHRSAESVFSG
ELQQLQPQRKCSHCGVQKTPQWRAGPMGAKTLCNACGVRYKSGRLLPEYRPACSPFTSSE
LHSNHRKVIEMRRKPTSDNETGLNQLVQSPQAVPSF*

>G1492 (149..919)
AATCCCAACCCACACACTCTCAAACTCCTCCTCTCCTCGTTTTCTTCTCTCTCCTCTTCA
CAGAACCACCAACATATCAAACTTTTTTCTCTTGGGTTTAAGTAAAAATCGAATCTTTG
TGTCGGTTTTTAGGGTTCTTGAAACGATATGGGTAAGTCTAGTGGTAGAAATGGTAACGG
AAGCTTTAACGGCAATAAATTTACGGAGTTAGACCTTACGTACGGTCTCCAGTTCCACG
GCTTAGATGGACGCCGGATCTTACCCTGTTGTTTCGTTACGCCGTCGAGATTCTCGGTGG
TCAACACCCGAGCAACACCAAACTTGTTCTTAAGATGATGGATGTGAAGGGACTTACCAT
TTCACATGTCAAAAGCCACCTTCAGATGTATAGAGGAGGTTCAAAGCTCACTTTGGAGAA
ACCAGAAGAAAGCTCATCATCTTCAATAAGAAGAAGACAAGACAGTGAAGAAGATTATTA
TCTTCATGACAACTTGCTTTTACACACAAGGAATGATTGCTTTTTGGGTTTTCACTCTTT
TCCTCTTTCTTACATTCTTCATTTAGAGGAGGAGGAGGAGGAAGAACAAGAGCAGCA
GACTTCAGAGTCTGGTGGTTATGATGATGATGCTGACTTTCTTACATCAAGAAGATGAA
CGATACGACGACGTTTTTGTACATCATTTCCCCAAGGGAACAGAGGAGTGGCGGGAACA
AGAACACGAAGAAGAAGAAGAAGATTTGTCGTTGCTCTGTGCTTAAATCATCATATTG
GAGAAGCAATGGATCATCGGTGGTGAGCGAAACGAGTGAAGCAGCAGTCTCGACTTGTTTC
AGCACCATTTCGTATCCAAAGATTGCTTTGGTTCTTCAAAGATTGATCTTAATCTGTCAAT
TTCTCTCTCGGTAGCTAAATAAGTTATGCAAGATTAGGTTTCAGAGAACTATTCCGAT
GTGTTTTTTGAACTAGGATATTGAATGTTAGTAGAGAAACCTAGAAAATGAAGTTTAGAT
AAATTATCAACGCAGCGTTTTGATCGCCTTTGAACGGAATAAACA

>G1492 Amino Acid Sequence (domain in AA coordinates: 34-83)
MGKSSGRNGNSFNGNKFHGVPRPVRSVPRLRWTPLHRCFVHAVEILGGQHRATPKLV
LKMDVKGTLTISHVKSHLQMYRGGSKLTLEKPESSSSSIRRRQDSEEDYLLHDNLSLHT
RNDCLLGFSPLSSHSSFRGGGGRTKEQQTSESAGGYDDADFLHIKMNDDTTTFLSHH

FPKGTEEWREQEHEEEEEEDLSLSLSLNHHHWSRNGSSVVSETSEAAVSTCSAPFVSKDCF
GSSKIDLNLNLSISLLGS*

>G1531 (1..666)

ATGTGTGAGTCAAGCAACAAAGTCAGAGTATCGCCATACCCGCTTCGGTCTTCGAGGACC
GACAAACACAAGGCGTCAGAGTCGCCTATTGAGACAGGTTGGGAGGATGTGCGTGATGT
CATCCTTACATGTGCGATACGAGTGTTCTGCTACTCCAATTGTTTCAAGCAGTTCGCGAGA
AAAACCATAAAAAGCGCCTATACCCCAAGACCTTACATTGTCCTCTCTGTAGAGGTGAA
GTATCCGAGACGACAAAGGTGACGAGCACTGCAAGAAGATTTATGAATGCTAAACCGAGG
TCTTGCTCCGTAGAGGATTGCAAATTCTCTGGGACGTTTTCTCAGCTTACTAAGCACTTG
AAAACCTGAGCATCGCGGTATTGTGCCACCAAGGTCGATCCACTGAGACAACAGAGATGG
GAAATGATGGAGAGACATTCTGAATACGTTGAACTCATGACTGCAGCTGGGATTTTCGCGT
ATGGCTGAGGTGATGCAACAACAGCTTCCCCAGGATCAGAATCATCTCATGTGTTTCAA
GTGACCGTTAATGGAACCATATGGAATCTAATTGATCCGAGTCAGGGAAGGAATGGATTA
GGCATCACCAACTATAGCGCAATGCAGTTTGTACCATTAAAGCATAAATCACAGTAGAACT
CTGTGA

>G1531 Amino Acid Sequence (domain in AA coordinates: 41-77)

MCESSNKVRVSPYPLRSSRTDKHKASESPIETGWEDVRGCHPYMCDTSVRHSNCFKQFRR
KTIKKRLYPKTLHCPLCRGEVSETTKVTSTARRFMNAKPRSCSVEDCKFSGTFSQLTKHL
KTEHRGIVPPKVDPLRQQRWEMMERHSEYVELMTAAGISRMAEVMQQQLPQDQNHVVFQ
VTVNGTIWNLIDPSQGRNGLGITNYSAMQFVPLSINHSRTL*

>G1540 (122..997)

atctctttactaccagcaagttgttttcttgctaacttcaaacttctctttctcttggtc
ctctctaagtcttgatcttatttaccgttaactttgtgaacaaaagtcgaatcaaacaca
catggagccgccacagcatcagcatcatcatcaagccgaccaagaaagcgggaacaa
caacaacaagtcgggtctgtgtggttacacgtgtcgccagaccagcagaggtggacacc
gacgacggagcaaatcaaaatcctcaaagaactttactacaacaatgcaatccgggtcacc
aacagccgatcagatccagaagatcactgcaaggctgagacagttcggaagattgaggg
caagaacgtcttttactggttccagaaccataaggctcgtgagcgtcagaagaagagatt
caacggaacaaacatgaccacaccatcttcatcaccacactcggttatgatggcgggctaa
cgatcattatcatcctctacttcaccatcatcacggtgttcccatgcagagacctgctaa
ttccgtcaacggttaacttaaccaagaccatcatctctatcatcataacaagccatatcc
cagcttcaataacgggaatttaaatcatgcaagctcagggtactgaatgtggtgttgtaa
tgcttctaattggctacatgagtagccatgtctatggatctatggaacaagactgttctat
gaattacaacaacgttaggtggaggtgggcaacatggatcatcattactcatctgcacc
ttacaacttctcgatagagcaaaagcctctgtttggtctagaaggtcatcaagacgaaga
agaatgtggtggcgatgcttctggaacatcgacgtacgcttctctcttccctatgca
cgggtgaagatcacatcaacgggtggttagtggtgccatctggaagtatggccaatcggaagt
tcgcccttgcgcttctcttgagctacgtctgaactagctcttacgccggtgtcgctcggg
attaaagctcttct
tagtgattaatgatgcagttgttatattagtagttaactagttatctctcgttatgtgta
atgtgtaattactagctaagtatcgtctaggtttaattgtaattgacaaccggttatctc
tatgatgaataagttaaatttatatat

>G1540 Amino Acid Sequence (domain in AA coordinates: 35-98)

MEPPQHQQHHHQADQESGNNNNKSGSGGYTCRQTSTRWTPTEQIKILKELYNNNAIRSP
TADQIQKITARLRQFGKIEGKNVFWFQNHKARERQKKRFNGTNMTTPSSSPNSVMMAN
DHYHPLLHHHHGVPMQRPANSVNVKLNQDHHLYHHNKPYPSPFNNGNLNHASSGTECGVVN
ASNGYMSHVYGSMEQDCSMNYNNVGGGWANMDHHYSSAPYNFFDRAKPLFGLGEGHQDEE
ECGGDAYLEHRRTLPLFPMHGEDHINGSGAIWKYQSEVRPCASLELRN*

>G1544 (1..2178)

ATGTCTCAGTCAAACATGGTACCAGTGGCTAACAAACGGAGACAACAACGACAACGAA
AACAAACAACAACAACAACAAATGGTGGAACTGACAACACTAATGCTGGAATGATTCT
GGAGATCAAGATTTTCGACAGTGGGAATACCTCAAGTGGCAATCATGGAGAAGGGTTGGGA
AACAAATCAAGCTCCTCGTCATAAGAAGAAAAAATACAATCGTCACACCCAACTTCAGATT
TCGGAGATGGAAGCTTTCTTCAGAGAGTGTCTCACCAGATGACAACAAAGGTACGAC
CTTAGCGCTCAATTGGGATTGGACCCTGTTTCAGATCAAATTCTGGTTCCAGAACAACGC
ACTCAAAACAAGAATCAACAAGAACGCTTTGAGAACTCAGAACTTCGGAATCTGAACAAC
CACCTTAGGTCTGAAAATCAGCGGTTACGAGAAGCTATTTCATCAAGCCTTATGCCCTAAG

TGTGGAGGCCAAACTGCAATTGGCGAAATGACCTTCGAAGAGCACCATCTTCGCATCCTC
AACGCTCGTTTACTGAAGAGATCAAGCAACTTTCCGTGACAGCGGAAAAGATATCAAGG
CTTACGGGGATACCAGTAAGGAGCCATCCCCGTGTGTCTCCTCCTAATCCTCCTCAAAT
TTCGAGTTCGGGATGGGATCTAAGGGAAATGTCGGAAACCACTCGAGGGAAACCACTGGA
CCTGCAGATGCTAATACCAAGCCGATCATCATGGAGTTGGCATTGAGCCATGGAGGAG
CTCTTGGTGTAGGCTCAAGTGGCTGAACCACTGTGGATGGGAGGATTTAATGGCACTAGC
TTAGCTTTGAACTTGGATGAATACGAAAAGACGTTTCGCACGGGTCTCGGTCTTAGACTT
GGCGGGTTTCGAACCGAGGCATCCAGGGAAACTGCACTCGTGGCAATGTGTCTACTGGC
ATTGTTGAAATGCTCATGCAAGAGAATCTGTGGTCAACAATGTTTGCCGGAATTGTTGGT
AGAGCCAGGACTCATGAACAGATAATGGCTGATGCTGCTGGAACCTTCAATGGAAATCTC
CAATAATGAGTGTGAGTACCAAGTGCTTTCCCGCTAGTCACAACCCGCGAAAGCTAC
TTCGTCCGCTACTGTAAGCAACAAGGAGAGGGTTGTGGGCGGTGGTTCGATATTTCCATC
GACCATCTCTCCCAAACATCAACCTAAAATGTCGCCGCCGACCTCTGGATGTCTGATT
CAAGAAATGCATAGTGGTTACTCCAAGGTTACATGGGTGGAACATGTGGAAGTAGATGAT
GCAGGAAGTTACAGCATCTTTGAGAAATTAATCTGTACTGGTCAAGCTTTTGCTGCTAAC
CGCTGGGTTGGTACATTGGTACGCCAGTGTGAGCGGATATCTAGCATCTTGTGCACAGAT
TTTCAATCTGTGATTCCGGTGATCACATAACGCTAACTAACCATGGAAGATGAGCATG
CTGAAGATAGCTGAGCGGATTGCGAGAACCTTCTTTGCTGGAATGACCAATGCGACGGGG
TCTACAAATATTTTCTGGTGTTGAAGGAGAAGATATCAGAGTGATGACAAATGAAGAGCGTG
AATGATCCAGGAAAGCCTCCCGGTGTCATTATTTGTGCAGCCACTTCTTTTGGCTTCTCT
GCTCTCTTAACACTGTCTTTGACTTCTCAGAGAGGCTACTCACCGACACAATTGGGAT
GTTCTCTGCAACGGAGAGATGATGCACAAGATAGCAGAGATTACGAATGGGATAGACAAA
AGGAACTGTGCAAGTTTACTCCGGCATGGACACACTAGCAAGAGCAAGATGATGATAGTT
CAAGAGACTTCTACTGACCCAACAGCTTCATTTGTGCTTTATGCGCCTGTTGATATGACA
TCAATGGATATTACTCTCCATGGAGGTGGTGATCCTGACTTTGTGGTGATCCTGCCTTCT
GGTTTTGTCTATTTTCCAGATGGTACGGGTAAGCCTGGAGGAAAAGAAGGAGGATCACTT
TTGACCAATTTCTTCCAAATGCTGGTTGAGTCAGGTCTGAGGCTAGGCTGAGTGTTAGC
TCTGTTGCAACTACTGAGAATCTGATTCTGATCAACCGTGCGGAGGATCAAAGATTTGTTT
CCTGTGCTAGACTGCTTGA

>G1544 Amino Acid Sequence (domain in AA coordinates: 64-124)

MSQSNMVPVANNGDNNNDNENNNNNNNNGGTDNTNAGNDSGDQDFDSGNTSSGNHGEGLG
NNQAPRHKKKKYNRHTQLQISEMEAFFRECPHPDDKQRYDLSAQLGLDPVQIKFWFQNK
TQNKNQERFENSELRNLNNHLRSENQRLREAIHQALCPKCGQTAIGEMTFEEHHLRIL
NARLTEEIKQLSVTAEKISRLTGIPVRSHPRVSPNPPNPFEGMGSKGNVGNHSRETTG
PADANTKPIIMELAFGAMEELLVMAQVAEPLWMGGFNGTSLALNLDEYEKTFRTGLGPRL
GGFRTEASRETALVAMCPTGIVEMLMQENLWSTMFAGIVGRARTHEQIMADAAGNFNGNL
QIMSAEYQVLSPLVTTRESYFVRYCKQQEGELWAVVDISIDHLLPNINLKRRRPSGCLI
QEMHSGYSKVTWVEHVEVDDAGSYSIFEKLICTGQAFANRWVGTLVQRQERISSILSTD
FQSVDSGDHITLTNHGKMSMLKIAERIAARTFFAGMTNATGSTIFSGVEGEDIRVMTMKSV
NDPGKPPGVIIICAATSFWLPPAPNTVDFDLREATHRHNDVLCNGEMMHKIAEITNGIDK
RNCASLLRHGHTSKSKMMIVQETSTDPTASFVLYAPVDMTSMITLHGGGDPDFVVLPS
GFAIFPDGTGKPGGKEGSSLLTISFQMLVESGPEARLSVSSVATTENLIRTTVRRIKDLF
PCQTA*

>G156 (39..755)

AGGAAGAGGGAGCCACTCATAAGAGGAAGAAGAGAGAGATGGGTAGAGGGAAGATAGAGA
TAAAGAAGATAGAGAATCAGACGGCGAGGCAAGTGACCTTCTCCAAGAGAAGAACTGGTC
TTATAAAGAAGACTCGTGAGCTCTCTATTCTCTGTGACGCTCACATCGGTCTCATCGTCT
TCTCAGCCACCGGAAAGCTTTCCGAGTTCTGCTCCGAACAGAACAGGATGCCTCAACTCA
TTGACCGATACTTGATACCAACGGATTGCGACTTCTGTATCATCATGACGACCAGGAGC
AATTGCACCATGAGATGGAACCTACTAAGAAGAGAGACATGTAACCTTGAGCTTCGTCTGC
GTCAATTCCATGGACATGACTTAGCCTCCATTCTCTTAATGAGCTTGACGGACTCGAGA
GACAGCTAGAACATTCTGTCTCAAGTCCGTGAGCGTAAGAGGAGGATGCTAGAAGAAG
ATAACAACAACATGTACCGTTGGCTTCATGAGCATCGTGCAGCGATGGAGTTTCAACAAG
CTGGGATAGATACCAAACAGGGGAGTATCAACAGTTTATAGAGCAGCTTCAGTGCTATA
AACCAGGGGAGTATCAGCAGTTTCTAGAGCAGCAGCAACAACAACCAACAGCGTTCTTC
AGCTTGCTACACTTCTTCTGAGATTGATCCTACTTACAATCTCCAGCTTGCTCAGCCTA
ATCTTCAAACGATCCAACGGCCCAGAATGATTAATACAATCTCAATAGATATCTACTC

>G1584 (160..1281)

>G1584 Amino Acid Sequence (domain in AA coordinates: TBD)

>G1587 (1..816)

84

TGTTACTACTACGAGTTCATGCCTCTGAAGAACTGA

>G1587 Amino Acid Sequence (conserved domain in AA coordinates:61-121)

MGYISNNMLINYLPLSTTQPPLLLTHCDINGNDHHQLITASSGEHDIDERKNNIPAAATL
RWNPTPEQITTTLEELYRSGTRTPTEQIQQIASKLRKYGRIEGKNVFWFQNHKARERLK
RRRREGGAIKPKHDKVDKSSSGGHRVDQTKLCPSPFPHNRPQPQHELDPASYNKDNANN
EDHGTTEESDQRASEVGKYATWRNLVWTSITQQPBEINIDENVNGEEBETRDNRTLNLFP
VREYQEKTRGLIEKTKACNYCYEYEFMPLKN*

>G1588 (1..2232)

ATGTACCATCCAAACATGTTTGAGAGCCATCATATGTTTCGATATGACCCCAAAGAGTACC
TCTGATAACGACTTGGGAATCACCGGTAGCCGAGAAGATGACTTTGAGACCAAGTCAGGT
ACCGAAGTCACTACTGAGAATCCTTCTGGTGAAGAGCTTCAAGATCCTAGCCAACGTCCC
AACAAAAAGAAGCGTTACCATCGCCACACGCAACGCCAAATTCAAGAGCTCGAATCATT
TTTAAGGAATGTCTCATCCAGATGATAAGCAACGAAAAGAGTTGAGCCGTGATCTCAAT
TTAGAGCCTCTTCAAGTTAAGTTTGGTTCCAAAACAAACGCACACAGATGAAGGCACAA
AGTGAGAGGCATGAGAACCAGATTCTAAAGTCAGACAATGACAAGCTCAGAGCAGAGAAC
AATAGATACAAAGAAGCTCTAAGCAATGCTACATGCCCTAACTGTGGCGGTCCAGCTGCT
ATTGGAGAAATGTCTTTTGACGAACAACATCTCAGGATCGAAAATGCTCGGCTCCGCGAA
GAGATTGATAGGATCTCTGCTATTGCTGCGAAATACGTTGGGAAGCCGTTAGGATCGTCT
TTCGCTCCACTAGCGATCCACGCGCTTCTCGTTCGCTTGATCTTGAAGTTGGAACTTT
GGGAACCAGACAGGCTTTGTAGGAGAAATGTATGGAACAGGGGACATTTTGAGGTCAGTT
TCGATTCCTTCTGAGACTGATAAGCCTATAATCGTGGAGCTAGCGGTTGCAGCTATGGAG
GAACTCGTGAGAAATGGCTCAAACCTGAGATCCTTTATGGCTTTCAACCGATAATTAGTC
GAGATTCTCAACGAAGAAGAGTATTTGAGAACGTTTCCGAGAGGAATTGGACCAAAGCCA
TTAGGATTAAGATCAGAGGCGTCAAGACAATCTGCAGTTGTTATAATGAATCACATCAAT
CTCGTTGAGATTCTCATGGATGTGAATCAATGGTCTTGTGTTTCTCTGGGATTGTGTCA
AGAGCCTTGACACTTGAAGTTCTTTCAACTGGAGTTGCTGGGAACATAACGGTGCTTTA
CAAGTGATGACAGCTGAGTTTCAAGTTCCATCACCCCTAGTCCCAACGCGTGAGAACTAC
TTTGAGATACTGCAAAACAACAGTGACGGCTCTTGGGCTGTGGTTGATGTCTCTTTG
GACAGCCTTAGACCAAGTACTCCAATCTTAAGAACTAGAAGAAGGCCCTTCAGGTTGTCTG
ATTCAGAATTTGCCCTAATGGTTATTCTAAGGTTACATGGATAGAGCATATGGAGGTAGAT
GATAGATCAGTTCACACATGTATAAACCGTTGGTTGAGTCCGGTTAGCTTTTCGGTGCG
AAACGTTGGGTGGCTACACTCGAACGACAATGCGAGCGGCTTGCTAGCTCCATGGCCAGC
AACATTCCTGGTGATCTTTCGGTGATAACGAGTCTTGAAGGAAGGAAGAGTATGTTGAAG
CTAGCTGAGAGAATGTTATGAGTTTCTGCAGTGGTGTGGCGCGTGCAGTGCACACGCT
TGGACAACAATGTCGACAACAGGATCCGATGATGTTCCGGTCATGACCCGCAAGAGTATG
GATGATCCAGGAAGACCTCCGGGTATTGTTCTTAGTGCAGTACTTCATTCTGGATCCCA
GTTGCTCCCAAACGTGTTTTGATTTCTTCCGTGACGAAAATTCAAGAAAAGAGTGGGAT
ATTCTGTCAAATGGAGGTATGGTTGAGGAAATGGCTCATATAGCCAATGGTCATGAACCT
GGAAACTGTGTCTCCTTGCTCCGAGTCAATAGTGGAACTCGAGCCAGAGCAACATGTTG
ATTCTACAAGAGAGCTGTACAGATGCATCAGGATCGTATGTGATTTACGCGCCAGTGGAT
ATAGTGGCGATGAATGTGGTTCTAAGCGGTGGAGATCCTGATTACGTGGCGTGTGTGCCG
TCTGGTTTGTGCTATTTTACCGGATGGTTCCGTTGGAGGAGGAGATGGGAATCAGCATCAG
GAAATGGTTTCTACTACTTCTTCTGGGAGTTGTGGTGGTTCGCTTTTAAACGTTGCGTTT
CAGATTCTTGTGACTCTGTTCTACAGCTAACTCTCACTTGGCTCGGTGGCTACGGTT
AATAGTCTGATCAAATGTACGGTGGAGAGGATTAAAGCTGCTGTTTCTGTGATGTTGGA
GGAGGAGCGTAG

>G1588 Amino Acid Sequence (domain in AA coordinates: 66-124)

MYHPNMFESHMFDMTPKSTSDNDLGITGSREDDFETKSGTEVTENPSSGEELQDPSQRP
NKKKRYHRHTQRQIQELESFFKECPHPDDKQRKELSRDLNLEPLQVKFWFQNKRTQMKAO
SERHENQILKSDNDKLAENNRKYEALSNATCPNCGGPAAIGEMSFDEQHLRIENARLRE
EIDRISAIKAYVKGKPLGSSFAPLAIHAPSRSLDLEVGNFNGQTGFVGEVMTGDIILRSV
SIPSETDKPIIVELAVAAMEELVRMAQTGDPLWLSTDNSVEILNEEYFRTFPRGIGPKP
LGLRSEASRQSAVVMNHNILVEILMDVNQWSCVFSGIVSRALTLVLSTGVAGNYNGAL
QVMTAEFQVPSPLVPTRENYFVRYCKQHS DGSAVVDVSLDSLRLPSTPILRTRRRPSCGL
IQELPNGYSKVTWIEHMEVDDRSVHNMYKPLVQSGLAFGAKRWATLERQCERLASSMAS
NIPGDL SVITSPEGRKSM LKLAERMVMSFCSGVGASTAHAWTTMSTTGSDDVVMTRKSM
DDPGRPPGIVLSAATSFWIPVAPKRVDFLRDENS RKEWDILSNGGMVQEMAHIANGHEP

GNCVSLLRVNSGNSQSNMLILQESCTDASGSYVIYAPVDIVAMNVVLSGGDPDYVALLP
SGFAILPDGSVGGGDNQHQEMVSTTSSGSCGSLLTVAFAQILVDSVPTAKLSLGSVATV
NSLIKCTVERIKA AVSCDVGGGA*

>G1589 (179..2221)

ACCAAATCACAATAGCAATCACACACATCTCCACAAACACAGCTTGAGATGATCATGAAA
CACGTGCATCCTCAGATCTCTATCAATCCAGCTTGGTGAAAGAAGGTCAAGAATTGAAAG
AGAATCAAAGAAAACGACGTCGTTTCATTCTGTGTAACTACTAATTATACATAGAT
GGCTGCTTACTTTTACGGAACCCACCGGAGATCTCTGCCGATCCGACGCTGGTCTTCA
AACGTTGATCCTCATGAATCCAACTACTTACGTTTACGTACACCCAACAAGACAACGACTC
GAACAACAACAACAGCAACAATAGCAACAACAACAACAACAACAACAACAACAACA
CAACAACAGTAGTTTTCGTTTTCTCGATTCCACGCGCGCAGCCAAACGCGAGCCAGCA
GTTTCGTCGGAATACCACTCTCAGGTCACGAAGCTGCTTCCATTACAGCCGCGACAACAT
CTCCGTACTTTCAGCGTTATCCTCCGCGGTGCAGTACAGTCTCTACGGTAGCCACCAAGT
GGATCCCACTCAGCAGCAAGCCGCGTGTGAGACGCCACGCGCGCAGCAAGGCCTCTCTTT
AACCCTCTCGTCTCAACAGCAGCAGCAACAGCAACATCATCAACAACACCAGCCTATTCA
CGTCGGATTCCGGTCCGGACATGGAGAAGATATCCGGTTCGGGTCTGGCTCTACAGGATC
GGGGGTAACAACCGGTATAGCTAATCTTGTAGCTCCAAGTACTTGAAGGCAGCACAAGA
GCTTCTTGACGAAGTAGTCAACGCTGATTCCGATGACATGAACGCTAAATCCCACTATT
CTCATCGAAAAAGGGTAGTTGCGGAAATGATAAACCTGTGCGAGAATCATCGGCCGCGC
TGGAGGAGAAGGTTCCGGTGGCGGAGCAGAAGCAGCCGGGAAACGTCCGGTGGAGCTAGG
CAGCGCAGAGAGACAAGAAATACAGATGAAGAAAGCAAACTTAGTAACATGCTTCATGA
GGTGGAGCAGAGATATAGACAGTACCACCAGCAGATGCAGATGGTGATCTCTTCGTTTCA
GCAAGCGGCAGGGATAGGATCAGCGAAGTCATACACGTCGCTAGCATTGAAAACCATATC
AAGACAGTTCGGTGTCTGAAAGAGGCGATCGCTGGTCAGATAAAAGCGGCCAACAAGAG
TCTTGGGGAGGAAGATTCAAGTGTCTGGTGTGGGAGGTTTGGAGGGTTCGAGGCTCAAGTT
CGTGGACCACCACTTGAGACAGCAAGAGCTCTTCAACAACCTGGGAATGATTCAACATCC
TTCCAATAATGCTTGGAGACCTCAACGTGGTCTCCAGAACGAGCCGTCTCAGTCTCCG
TGCTTGGCTCTTCGAACACTTTCTTCATCCATACCCTAAGGATTTCGGACAAGCACATGCT
AGCTAAGCAAACAGGACTCACTCGTAGCCAGGTGTGCAACTGGTTTTATAACGCGAGAGT
TCGGTTATGGAACCAATGGTGGAGGAGATGTACATGGAGGAAATGAAGGAGCAGGCAAA
GAACATGGGATCCATGGAAGAACTCCTTTGGATCAAAGCAACGAAGATTCTGCTTCAA
GTCAACAAGTAACCAAGAAAAGAGCCCAATGGCGGACACTAATTACCATATGAATCCCAA
TCACAACGGTGACCTAGAAGGCGTCACTGGAATGCAAGGATGCCCAAGAGACTAAGAAC
CAGCGACGAGACAATGATGCAGCCAATAAATGCGGATTTACGCTCCAACGAGAAGCTCAC
GATGAAGATTCTAGAAGAACGGCAAGGGATAAGATCAGATGGTGGCTACCTTTTCATGGG
TAATTTCCGGCAATACCAAATGGATGAGATGTCAAGATTTGATGTAGTCTCAGACCAGGA
GCTCATGGCGCAAAAGTACTCAGGAAACAACAATGGCGTGTCCCTCACGTTAGGTTTACC
TCATTGTGATAGCTTGTCTGTCACGGACCATCAGGGTTTCATGCAGACCCACCATGGGAT
TCCTATAGGGAGAAGAGTGAAAATAGGAGAAACAGAGGAATATGGACCCGCCACCATCAA
TGGTGGTAGCTCGACCACAACCGCACATTTCATCAGCGGCAGCTGCCGCGGCTTACAATGG
GATGAACATACAGAACCAGAAGAGATATGTGGCTCAGTTATTGCCCCGACTTCGTTGCATA
AACCCATCTCTCTAGAAGGAGAAACCGAAACAGGTTATTATATACGTTTCTAGTTTTTAA
TTAGTATATAGTTTCTCATACCATTGAACCAAAACAAGAACAAAATTTAATTTTAGTCT
TTGGTTATATATGCGCGACGGGCTACGTCAGGGCCCTGACGTAGC

>G1589 Amino Acid Sequence (conserved domain in AA coordinates:384-448)

MAAYFHGNPPEISAGSDGGLQLTLILMNPITYVQYTQDNDNSNNNNNSNNNNNTNTNTN
NNNSSFVFLD SHAPQPNASQQFVGIPLSGHEAASITAADNISVLHGYP PRVQYSLYGSHQ
VDP THQQAACETPRAQQGLSLTLSSQQQQQQHHQHQPIHVGFSGHGEDI RVGSGSTG
SGVTNGIANLVSSKYLKAAQELLDEVVNADSDDMNAKSLFSSKKGSGCNDKPVGESSAG
AGGEGSGGGAEAGKRPVELGTAERQEI QMKKAKLSNMLHEVEQRYRQYHQMQMVISF
EQAAGIGSAKSYTSLALKTISRQFRCLKEA IAGQIKAANKSLGEEDSVSGVGRFEGSR LK
FVDHHLRQORALQQLGMIQHPSNNAWRPQRGLPERAVSVLRWLFEHFLHPYPKDS DKHM
LAKQTGLTRSQVSNWF INARVRLWKPMVEEMYMEEMKEQAKNMGSM EKTPLDQSNEDSAS
KSTSNQEKSPMADTNYHMPNHNGLDLEGVTGMQGPC KRLRTSDETMMPINAD FSSNEKL
TMKILEERQGISRDGGYPFMGNFGQYQMDMSRFDVVS DQELMAQRYSGNNNGVSLTLGL
PHCDLSSTDHQGF MQTHHGIPIGRRVKIGETEBYGPATINGGSSTTAHSSAAAAAAYN
GMNIQNQKRYVAQLLPDFVA*

>G160 (38..784)

TCAAATTTGTCATTTGTTTATTCAAATTTTTGAGAAAATGGTGAGAAGTACCAAAGGTCG
TCAGAAAATAGAGATGAAAAAATGGAACGAAAGCAACCTTCAGGTTACTTTCTCAA
AAGAAGATTCCGCTCTTTTCAAAAAGCTAGTGAACTTTGCACATTAAGTGGTGACAGAT
TCTGTTGATTGTGTTCTCTCCTGGTGGGAAAGTGTCTTTTGGCCATCCAAGTGTTC
AGAAGTCAATTCATCGCTTTTTCGAATCCTAACATAATTCTGCCATTGTCCATCATCAGAA
CAACAATCTCCAACCTGTTGAAACCCGTCGGATAGAAATATCCAATATCTCAACAATAT
ACTCACTGAGGTGCTGGCAAACAGGAAAAGGAGAAACAGAAGAGAATGGTTTGGACCT
ATTGAAAGAATCCAGAGAACAACTAGGAAACTGGTATGAAAAAGATGTGAAAGATCTCGA
CATGAATGAAACCAACCAGCTGATATCTGCTCTTCAAGATGTGAAAAAGAACTGGTAAG
AGAAATGTCTCAATATTCTCAAGTAAATGTTTCGCAGAATTACTTTGGTCAAAGTTCTGG
CGTGATTGGTGGTGGTAATGTTGGCATTGATCTTTTGGATCAAAGAAGAAATGCATTCAA
CTATAATCCAAACATGGTGTCTTCCCAATCATACACCACCAATGTTTGGATACAACAATGA
TGGAGTTCTCGTTCCGATATCCAACATGAAGTACATGTCAAGTTACAACCTCAACCAGAG
CTAGAGTCTGAAGCTAGAAGAATCCTAATCAATATTTGCGTTATTTTGGCTATGGTTA
CTGTTAGGATTGTTCTTGTATTGTGAGACTTAAGTTTGTCTTTTCTTTAATTGTTTCA
GTTGGTTGGTTTTCATTTTATTCTGCTGTTTGTCTTTTCTTTTGGATATTTTGTGTA
TCCCAGAATAAATTTATTTATCCTTTAAAAA

>G160 Amino Acid Sequence (domain in AA coordinates: 7-62)

MVRSTKGRQKIEKMKMENESNLQVTFSKRRFGLFKKASELCTLSGAEILLIVFSPGGKVF
SFGHPSVQELIHRFSNPNHNSAIVHHQNNNLQLVETRPDRNIQYLNILTEVLANQEKEK
QKRMVLDLLKESREQVGNWYEKDVKDLDMNETNQLISALQDVKKLVREMSQYSQVNVSQ
NYFQSSGVIGGNGVIGIDLFDQRRNAFNYPNMFVFNHTPPMFGYNNDGVLVPIINMNYM
SSYNFNQS*

>G1636 (19..666)

GAGTAATCATCAACGATTATGGCGTCAAGTCAGTGGACGAGGTCCGAGGATAAGATGTTT
GAGCAAGCTTTGGTTCTTTTTCTGAAGGATCTCCTAATCGGTGGGAGAGAATCGCTGAT
CAGCTTCATAAATCTGCTGGTGAAGTTAGGGAGCATTACGAGGTCTTGGTTCATGATGTT
TTCGAGATTGATCTCGGTGAGTTGATGTCCCTGATTACATGGATGACTCGGCGCTGCG
GCGGCGGGTTGGGATTCGCTGGTCAGATCTCTTTTGGGTCTAAACATGGCGAGAGTGAA
CGCAAAAGAGGAACTCCTTGGACAGAGAACGAACACAAATGTTTCTGATCGGATTAAAG
AGATATGGTAAGGGAGATTGGAGGAGTATCTCGAGAAACGTTGTGGTGACGAGGACACCG
ACGCAAGTCGCGAGTCACGCTCAGAAGTATTTTCTGAGACAGAACTCGGTGAAGAGGAG
AGGAAAAGGTCGAGCATCCATGATATAACTACGGTTGATGCTACTTTGGCTATGCCTGGG
TCTAACATGGACTGGACTGGCCAACACGGGAGTCCTGTTCAAGCGCCGAGCAGCAACAG
ATTATGCTCTGAGTTCGGTCAGCAATTGAATCCTGGTCATTCGAGGATTTTGGGTTTCGG
ATGTGATG

>G1636 Amino Acid Sequence (domain in AA coordinates: 100-165)

MASSQWTRSEDKMFEQALVLFPEGSPNRWERIADQLHKSAGEVREHYEVLVHDFEIDSG
RVDVPDYMDSSAAAAGWDSAGQISFGSKHGESERKRGPWTENEHKLFLIGLKRYGKGD
WRSISRNVVTRTPQTQVASHAQKYFLRQNSVKKERKRSSIHDIITVDATLAMPGSNMDWT
GQHGPSVQAPQQQIMSEFGQLNPGHFEDFGFRM*

>G1642 (1..1077)

ATGGGTCATCACTCATGCTGCAACAAGCAAAAGGTGAAGAGAGGGCTTTGGTCACCTGAA
GAAGACGAAAAGCTCATCAACTACATCAATTCATATGGCCATGGATGTTGGAGCTCTGTT
CCTAAACATGCAGGTTTGCAGAGATGTGGAAAGAGTTGTAGATTAAGATGGATAAATTAT
CTAAGACCTGATCTTAAACGTGGAAGCTTCTCTCCTCAAGAAGCTGCTCTTATCATTGAG
CTTCACAGCATTCTTGGTAACAGATGGGCTCAAATTGCTAAACATCTACCTGGAAGAACA
GATAACGAGGTCAAGAATTTCTGGAACCTCGAGCATTAAAAAGAAGCTCATGTCTCACCAT
CATCACGGTCATCATCATCATCTCTCTTCCATGGCGAGTTTGCTCACAAACCTTCTCT
TATCACAATGGGATTCACCCCTACTACAGTCGACGATGAAAGTTCAAGATTATGTCCAAT
ATCATCAACAACACTAACCTAATTTTCATCACTCCAAGCCATCTCTCTCTTCTCTCTCTCT
CATGTTATGACCCCATGATGTTGCCAACCTCTAGAGAAGGAGATTTCAAGTTTCTAACC
ACAAACAACCCAAACCAATCTCATCACCATGATAATAACCATTACAACAACCTCGACATT
TTGTCAACCCACACCAACTATAACAATCATCATCAACCTTCACTTTCTTCTTGTCTCAT
GATAATAATCTCCAATGGCCAGCGTTACCAGATTTCCAGCGAGTACCATTTCTGGTTTC
CAAGAAACCCCTCAAGATTATGATGATGCTAATAAACTCAACGTGTTTGTGACACCATTC

AACGATAATGCCAAAAAGTTATTATGTGGAGAAGTTCTCGAAGGCAAAGTACTATCTTCC
TCCTCACCAATTTTACAAGATCACGGCCTTTTTCTTCCCACCACGTACAACCTTTCAAATG
ACTTCTACGAGTGATCATCAACATCATCATCGAGTGGACTCATACATCAATCACATGATC
ATACCATCATCATCCTCATCGTCGCCAATCTCTTGTGGACAGTACGTCATAACTTAA

>G1642 Amino Acid Sequence (domain in AA coordinates: TBD)

MGHSHCCNKQKVKRGLWSPEEDEKLINYINSYGHGCWSSVPKHAGLQRCGKSCRLRWINY
LRPDLKRGFSFSPQEAALIIELHSILGNRWAQIAKHLPGRTDNEVKNFWNSSIKKLMSSH
HHGHHHHHLLSSMASLLTNLPYHNGFNPTTVDDESSRFMSNIIITNPNFITPSHLSLPSP
HVMTPLMFPPTSREGDFKFLTNNPNQSHHHDNNHYNLDILSPTPTINNHHQPSLSSCPH
DNNLQWPALPDFPASTISGFQETLQDYDDANKLVFVTPFNDNAKLLCGEVLEGKVLSS
SSPISQDHGLFLPTTYNFQMTSTSDHQHHRVDSYINHMIIIPSSSSSSPISCGQYVIT*

>G1747 (1..777)

ATGAAAAATGATGCAAGAGGAGGAAACCGAAAAGGTCCATGGACAGAACAGGAAGACATA
CTTCTGGTAAATTTTGTTCACCTTATTTGGAGATCGACGATGGGATTTTATAGCAAAAGTA
TCAGGTTTGAACAGAACAGGAAAGAGTTGCAGGCTAAGATGGGTTAATTACCTACATCCT
GGTCTCAAACGTGGCAAGATGACGCCTCAAGAAGAGCGCCTCGTCCCTGAGCTTCACGCT
AAGTGGGGAAACAGGTGGTCGAAAATAGCCCCGAAAATTGCCGGGACGAACGGATAACGAG
ATAAAGAACTACTGGAGGACTCATATGAGGAAGAAAGCTCAAGAAAAGAAGCGTCTGT
TCCCCAACTTCTCATTTTCCAACCTGCAGCTCGTCATCTGTGACCACTACCACCACCAAT
ACTCAAGATACATCGTGCCACTCGCGTAAATCTTCAGGGGAAGTGAGCTTTTACGACACT
GGAGGTTCCCGATCCACTAGAGAGATGAATCAAGAAAACGAAGACGTGTACTCGTTGGAT
GATATATGGAGAGAGATTGATCACTCAGCAGTAAACATAATAAAACCGGTTAAAGACATC
TACTCAGAACAAAGCCATTGCTTAAGTTACCCAAATCTAGCTTCACCATCATGGGAAAGC
TCATTGGATTCTATATGGAACATGGATGCAGATAAAAGTAAGATATCGTCTTACTTTGCA
AATGATCAGTTTCTCTTCTGTTTCCAACACAGTAGATCACCATGGTTCGTGAGGTTAA

>G1747 Amino Acid Sequence (domain in AA coordinates: 11-114)

MKMMQEEGNRKGPWTEQEDILLVNFVHLFGDRRWDFAIKVSGLNRTGKSCRLRWVNYLHP
GLKRGKMTPOEERLVLELHAKWGNRWSKIARKLPGRDTDNEIKNYWRTHMRKKAQEKRPV
SPTSSFSNCSSTSVTTTTNTQDTSCHSRKSSGEVSFYDTGGSRTREMNQENEDVYSLD
DIWREIDHSAVNIIPVKDIYSEQSHCLSYPNLASPSWESSLDSIWNMDADKSKISSYFA
NDQFPFCFQHSRSPWSSG*

>G1749 (59..535)

CAACACTTCTCAGTGACCGTGAGCAACGAATTATTTTCAGTTCAACGACTCCGCGGAAAT
GGAAAATTCAGAAAATGTTCCCTCTTACGATCAAAACATCAATTTCACTCCTAATTTGAC
GAGAGATCAAGAACATGTGATCATGGTCTCTGCTTTGCAACAAGTAATATCCAACGTCGG
AGGTGACACGAACTCGAATGCATGGGAAGCTGATCTTCCACCTTTGAACGCTGGCCCTTG
TCCTCTTTGTAGTGTACCGGCTGCTACGGTTGCGTCTTCCCACGACACGAGGCGATAAT
TAAGAAGGAGAAGAACACAAAGGAGTGAGGAAAAAACCATCAGGTAAATGGGCGGCGGA
GATATGGGATCCGAGTTTGAAAGTAAGGAGATGGCTTGGAACGTTTCCAACAGCGGAGAT
GGCGGCTAAGGCTTACAACGATGCGGCGGCTGAGTTTGTGCGAAGAAGATCAGCAAGACG
TGGCACAAGAAGACGAGAGGAAGCATCTACCAAGAAGACGACTGAGAAAAATTAACGGAG
AAGGAGCACGTATAGAAAGGCAGGAAGAGGCATCTTACTTGCTTCAAGTAATCAGAA
TTTTTTTGAAAAGTAAAAACGTTATTTTGTGTTGGTAATAAAATAAAGTAAAAACAAATAT
TGCTAACGCAAGACTTATCAAGTTCAGTCGTGACTGTGAGTGTGTTTTATGTATCTTAC
TTCATTTTTTTGTCTTTCAATTGTGTGTGTGTGT

>G1749 Amino Acid Sequence (conserved domain in AA coordinates: 84-155)

MENSENVPYDQNIPTNLTDRDQEHVIMVSALQOVISNVGGDTNSNAWEADLPPLNAGP
CPLCSVTGCGYCVFPRHEAIIKKEKKHKGVRRKPSGKWAAEIWDPSLKVRRLGTFFPTAE
MAAKAYNDAAAEEFVGRRSARRGKNGEESTKKTTEKN*

>G1751 (117..923)

AAACACAAACAAAACATATTTTCAATCTCCAGGTGCTTTACACCAACAGAGTCGCAAG
AAAAACAAAACCAAACCTCGGATTTAGTTTGACAGAAGAAGGAATCGAGAGTCGGGTATGC
ATTATCCTAACACAGAACCAATTCGTCGAGCTCCAGCCCCAACCCGGTATCAAAAGG
AGCAGTTGTACCCGGAGCAAGAGCTTTCAGTTATTGTCTCTGCTTTGCAACACGTGATCT
CAGGGGAAAAACGAAACGGCGCGCTGTGAGGTTTTTCCAGTGACAGCACAGTGATAAGCG
CGGGAATGCCTCGGTTGGATTAGACACTTGTCAAGTCTGTAGGATCGAAGGATGTCTCG
GCTGTAACACTATTTTTTCGCCCAAATCAGAGAATTGAAAAGAATCATCAACAAGAAGAAG

AGATTACTAGTAGTAGTAACAGAAGAAGAGAGAGCTCTCCCGTGGCGAAGAAAGCGGAAG
GTGGCGGGAAAATCAGGAAGAGGAAGAACAAGAAGAATGGTTACAGAGGAGTTAGGCAAA
GACCTTGGGGAAAATTTGCAGCTGAGATCAGAGATCCTAAAAGAGCCACACGTGTTTGGC
TTGGTACTTTTCAAACCGCCGAAGATGCGGCTCGAGCTTATGATCGAGCCGCGATTGGAT
TCCGTGGGCCAAGGGCTAAACTCAACTTCCCCTTTGTGGATTACACGTCTTCAGTTTCAT
CTCCTGTTGCTGCTGATGATATAGGAGCAAAGGCAAGTGCAAGCGCCAGTGTGAGCGCCA
CAGATTTCAGTTGAAGCAGAGCAATGGAACGGAGGAGGAGGGGATTGCAATATGGAGGAGT
GGATGAATATGATGATGATGATGGATTTTGGGAATGGAGATTCTTCAGATTTCAGGAAATA
CAATTGCTGATATGTTCCAGTGATAAATGAGCTCTTTCTTGTGGCGTTTTTGGAGTTA
AGTGCAAGAAGAGATTGACACTGTGGCTTGTTTAAAGTGAACAAGAAACAAGAAAGCATGT
AATTAGTAGTCTCATTCTTTTGTGTTGTGGTCAATTCTATGTTTATCTCATATAAAATCTG
AGTTAAACCTATCTGAGGAGAGAGTAAATAAAGAGGTTAAGAA

>G1751 Amino Acid Sequence (domain in AA coordinates: TBD)
MHYPNNRTEFVGAPAPTRYQKEQLSPEQELSVIVSALQHVISGENETAPCQGFSSDSTVI
SAGMPRLDSDTQCVCRIEGLGCNYFFAPNQRIEKNHQEEEEITSSSNRRRESSPVAKKA
EGGGKIRKRKNKNGYRGVRQRPWGKFAAEIRDPKRATRVWLGTFTAEADAARAYDRAAI
GFRGPRAKLNFPFVDYTSVSSPVAADDIGAKASASVSATDSVEAEQWNGGGGDCNME
EWMNMMMMMDFGNGDSSDSGNTIADMFAQ*

>G1752 (25..756)
AAAAAAAAAAAAAAAAAACTTATGGAATATTCCCAATCTTCCATGTATTCATCTCCA
AGTTCTTGAGCTCATCACAAGAAATCACTCTTATGGAACGAGAGCTGTTTCTTGATCAA
TCATCTGAACCTCAAGCCTTCTTTGCCCTAATTATGATTACTCCGATGACTTTTTCTCA
TTTGAGTCACCGGAGATGATGATTAAGGAAGAAATTCAAACGGCGACGTTTCTAACTCC
GAAGAAGAAGAAAAGGTTGGAATGATGAAGAAAGATCATAACAGAGGAGTGAGGAAAAGG
CCGTGGGGGAAATTTGCAGCGGAGATAAGAGATTCAACGAGGAATGGAATTAGGGTTTGG
CTCGGGACATTTGACAAAGCCGAGGAAGCCGCTCTTGCTTATGATCAAGCGGCTTTCGCC
ACAAAAGGATCTCTTGCAACACTTAATTTCCCGGTGGAAGTGGTTAGAGAGTCGCTAAAG
AAAATGGAGAATGTGAATCTTCATGATGGAGGATCTCCGGTTATGGCCTTGAAGAGAAAA
CATTCTCTTCGAAACCGGCCCTAGAGGGAAAAAGCGATCCTCTTCTCTTCTTCTTCTTCT
TCTAATTCTTCTTCTTGCTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT
AAGCAGAGTGTTGTGAAGCAAGAAAGTGGTACACTTGTGGTTTTTGAAGATTTAGGTGCT
GAGTATTTAGAACAACCTTCTTATGAGCTCATGTTGATCTTGTAAATTGATTTCAGCAAAAG
CCACTATTAACTTTAATTTTGTGATAATTAATCTTGAAATTTGTTTTGTTCTATTCTGCA
ATTTCTTTGGTTCTCTTATTTTTTGTGTTGTATCCAAATGAAATTATTGGAAGAGATG
GTGATGTTAAAGTGATATATATATAAAAAAAAAA

>G1752 Amino Acid Sequence (domain in AA coordinates: TBD)
MEYSQSSMYSSPSSWSSSQESLLWNESCFLDQSSEPQAFFCPNYDYSDDFSFESEPEMMI
KEEIQNGDVNSNEEBEKGVIDEERSYRGVRKRPWGKFAAEIRDSTRNGIRVWLGTFDKAE
EAAALAYDQAAFATKGSLATLNFPEVVRESLKKMENVNLDGGSPVMALKRKHSLRNRPR
GKKRSSSSSSSSSSSSSSSSSSSTSRSSSKQSVVKQESGTLVVFEDLGAEYLEQLLM
SSC*

>G1763 (33..977)
GTACATTTTTTTTGTATTTTCAGGAACTCCGATGGCGGATCTCTTCGGTGGTGGCCACG
GCGGCGAGCTTATGGAAGCACTTCAACCTTTTACAAAAGTGCTTCACGTCTGCTTCAA
ATCCTGCGTTTGCCTCCTCAAACGATGCGTTTGCCTCTGCCCCAAACGACCCATTTCTT
CTTCTTCTTACTATAATCCTCATGCATCTTTCTTCCCTTCACATTCCACAACCACTTACC
CGGATATTTATCTGGATCCATGACCTATCCATCTTCATTCCGGGTGGATCTTCAACAAC
CCGAAAACCTACCAATCTCAGTTCCATTACCAAAACACTATCACTTACACTACCAAGACA
ACAACACTTGATGCTCAACTTCATTGAGCCGAGCCAACCGGATTTTATGACCCAACCGG
GTCCGAGTTCCGGTTGCGGTTTCAAACCGGCTAAGCTCTATAGAGGAGTGAGGCAAGAC
ATTGGGGAAAATGGGTGCGGAGATCCGTTTACCCAGGAACCGAACCCGACTTTGGCTCG
GAACATTCGACACGGCTGAAGAAGCCGCTTGGCTTATGATCGCGCCGCTTTAAGCTTC
GTGGTGACTCGGCTCGGCTTAACCTCCAGCTCTCCGATACCAAAACCGGCTCGTCTCCGT
CTGACGTTGGCGAATACGGACCTATTCAAGCTGCGGTTGACGCCAAGCTAGAAGCCATAT
TAGCTGAGCCGAAGATCAGCCGGGCAAAACGGAGAGGACGTCGAGGAAACGAGCTAAAG
CCGCGGCTTCTTCAGCTGAGCAGCCGTCAGCGCCACAACAACATTCCGGGTGCGGTGAAA
GTGATGGGTGCGGTTACCGACTTCGGATGTTATGGTGCAGGAGATGTGCCAAGAGCCAG

AGATGCCATGGAATGAAAATTTTCATGCTCGGCAAGTGTCTTCTTATGAGATAGATTGGG
CTTCAATTTTATCGTGAATAATAGGATTCAATTCATTTTATTCAATTTAACTTGTTTG
TATTTTCTTTTAACTTTAGGGTTATTAGCTGTGCGTAAAATTTGTAAATTTAGCATTTTG
TATGAATGTAATGCAAGTGTGTAAATTATGGACAGCTCAAGCTTTTTTGTAAAA

>G1763 Amino Acid Sequence (conserved domain in AA coordinates:140-209)

MADLFGGGHGGELMEALQPFYKSASTSASNPFASSNDASFASAPNDPSSSSSYNPHASF
FPSHSTTTTPDIYSGSMTYPSSFGSDLQOPENYQSQFHYQNTITYTHQDNNTCMLNFI
EQPDMFTQPGPSSGSVSKPAKLYRGVQRHWGKWVAEIRLPRNRLWLGTFTABEAAAL
AYDRAAFKLRGDSARLNFALRYQTGSSPSDVGEYGPIQAAVDKLEAILAEPKNQPGKT
ERTSRKRAAAAASSAEQPSAPQQHSGSGESDGSPTSDDVMVQEMCQEPMPWNEFMLG
KCPSYEIDWASILS*

>G1766 (32..1216)

AGGCTATTCTCGAAAAACAAAGAATAAAGAATGAATTCGTTTTACAAGTACCTCCTGG
CTTCAGATTTTCATCCTACTGATGAAGAACTTGTAGACTACTACTTGAGGAAAAAGTTGC
ATCAAAGAGAATAGAAATCGATATCATCAAGGATGTTGATCTTTACAAGATTGAGCCATG
TGATCTTCAAGAGTTATGCAAGATAGGAAACGAAGAGCAGAGCGAATGGTACTTCTTTAG
TCATAAAGACAAGAAGTATCCACGGAACCTCGAACCAATAGAGCCACGAAAGCAGGATT
TTGGAAAGCCACTGGAAGAGACAAGGCTATATATATAAGACATAGTCTTATCGGTATGAG
GAAAACACTTGTGTTTTACAAAGGAAGAGCCCCAAATGGTCAGAAATCCGATTGGATCAT
GCACGAATATCGCTTAGAAACAAGTGAATGGAACCCCTCAGGAAGAAGGATGGGTAGT
ATGTAGGGTATTCAAGAAGAAATTTGGCAGCGACAGTGAGGAAATGGGAGATTACCATT
ATCACCATCGCAGCATTGGTACGATGATCAGCTCTCTTTTATGGCCTCCGAGATCATTT
TAGTTCTCCACGACAGTTTCTTCCCAATCATCATTATAACCGCCACCATCACCAGCAGAC
ATTGCCTTGTGGCTCAATGCATTCAACAACAACATCCTAATTGCAATGCAAGCAAGA
GCTCGAGTTACATTACAATCAAATGGTACAACATCAACAACAAAACCATCATCTTCGTGA
ATCTATGTTTCTCCAGCTTCTCAGCTCGAAAGCCCTACCAGTAATTGCAATTCTGACAA
CAACAATAACACAAGAAATATTAGTAACTTGCAAGAAATCATCAAATATATCTCATGAGGA
ACAATTGCAACAAGGAATCAAAGTTTCAGCTCTCTGTATTACGATCAAGGAGTAGAGCA
AATGACTACTGACTGGAGAGTTCTCGATAAAATTTGTTGCTTCACAGCTTAGCAATGATGA
AGAGGCTGCAGCCGTGGTTTTCTTCTTCTCATCAAAACAACGTCAAGATTGACACGAG
AAACACGGGTTATCATGTGATAGATGAGGGAATAAATTTGCCGAGAATGATTCTGAAAG
GGTTGTTGAAATGGGAGAAGAGTATTCAAATGCTCATGCTGCTTCTACTTCTTCAAGTTG
TCAGATTGATCTCTAGAAATAGTGATAGAGAGATGAAAAGATGCAAGGTGAATATATAT
GAAAATACATGCACACTAGTGTTATTTATACTTAAAGATGGAAGGGGAAAAACAAGGAGT
TATTTCTGGATTTATGGAGGTTTTGTACATAATAAAACCTACAACCATATGGTATTTT
CTTTTGAAAAA

>G1766 Amino Acid Sequence (domain in AA coordinates: 10-153)

MNSFSQVPPGFRHPTDEELVDYLRKKVASKRIEIDIIDVDLYKIEPCDLQELCKIGN
EEQSEWYFFSHKDKKPYTGRTNRATKAGFWKATGRDKAIYIRHSLIGMRKTLVFKGRA
PNGQKSDWIMHEYRLTSENGTPQEEGWVCRVFKKLAATVRKMGDYHSSPSQHWYDDQ
LSFMASEIISSSPRQFLPNHHYNRHHHQQTLPCGLNAFNNNNPNLQCKQLELHYNQMVQ
HQQQNHHLRESMFLQLPQLESPTSNCSNNDNMNTRNISNLQKSSNISHEEQLOQGNQSF
SLYYDQGVQMTDWRVLDKFVASQLSNDEEAAAVVSSSHQNNVKIDTRNTGYHVIDEG
INLPENDSERVEMGEEYSNAHAASTSSSCQIDL*

>G1767 (1..1596)

ATGGATACTCTCTTTAGACTAGTCAGTCTCCAACAACAACAATCCGATAGTATCATT
ACAAATCAATCTTCGTTAAGCAGAACTTCCACCACCCTACTGGCTCTCCACAACTGCT
TATCACTACAACCTTCCACAAAACGACGTCGTCGAAGAATGCTTCAACTTTTTTCATGGAT
GAAGAAGACCTTTCTCTTCTTCTTCTCACCACAACCATCACAACCACAACAATCCTAAT
ACTTACTACTCTCTTCTACTACTCCACCCAATACCATCCCGCCACATCATCAACCCCT
TCTTCCACCGCCGAGCCGAGCTTTAGCCTCGCCTTACTCTCTCTCCGGCCACCATAAT
GACCTTCCGCGTTCTCCATACCTCAAACCTCCTCGCTCCTTCACTTCTCAGCCAATGCC
AAGTGGGCAGACTCGGTCTCTTGAAGCGGCACGTGCCTTCTCCGACAAAGACACTGCA
CGTCCGCAACAATCCTATGGACGCTCAACGAGCTCTCTTCTCCGTACGGAGACACCGAG
CAAAAACCTGGCTTCTTACTTCTTCCAAGCTCTTCAACCGCATGACCGGTTTCAAGCGAA
CGATGCTACCGAACCATGGTAACAGCTGCAGCCACAGAGAAGACTTGCTCCTTCCAGTCA
ACGCGAAAACTGTACTAAAGTTCCAAGAAGTTAGCCCTGGGCCACGTTTGGACACGCTG

CGCGCAAACGGAGCAATCTTGAAGCAGTAGACGGAGAGGCAAAGATCCACATCGTTGAC
ATAAGCTCCACGTTTTCACCTCAATGGCCGACTCTTCTAGAAGCTTTAGCCACAAGATCA
GACGACACGCCTCACCTAAGGCTAACACAGTTGTCGTGGCCAACAAGTTTGTCAACGAT
CAAACGGCGTCGCATCGGATGATGAAAGAGATCGGAAACCGAATGGAGAAATTCGCTAGG
CTTATGGGAGTTCTTTTCAAATTTAACATTATTCATCACGTTGGAGATTATCTGAGTTT
GATCTCAACGAACTCGACGTTAAACCAGACGAAGTCTTGGCCATTAAGTGGTAGGCGCG
ATGCATGGGATCGCTTTCACGTGGAAGCCCTAGAGACGCTGTGATATCGAGTTTCCGACGG
TTAAGACCGAGGATTGTGACGGTCGTAGAAGAAGAAGCTGATCTTGTGCGAGAAGAAGAA
GGTGGCTTTGATGATGAGTTCTTGAGAGGGTTTGGAGAATGTTTACGATGGTTTAGGGTT
TGCTTCGAGTCATGGGAAGAGAGTTTCCAAGGACGAGCAACGAGAGGTTGATGCTAGAG
CGTGCAGCGGGACGTGCGATCGTTGATCTTGTGGCTTGTGAGCCGTCGGATTCCACGGAG
AGGCGAGAGACAGCGAGGAAGTGGTCGAGGAGGATGAGGAATAGTGGGTTTGGAGCGGTG
GGGTATAGTGATGAGGTGGCGGATGATGTCAGAGCTTTGTTGAGGAGATATAAAGAAGGT
GTTTGGTCGATGGTACAGTGTCTGATGCCGCCGGAATATTCCTTTGTTGGAGAGATCAG
CCGGTGGTTTGGGCTAGTGCCTGGCGGCCAACGTAA

>G1767 Amino Acid Sequence (domain in AA coordinates: 255-272)

MDTLFRLVSLQQQQSDSIIITNQSSLSRTSTTTTGSPTAYHYNFPQNDVVEECFNFMD
EEDLSSSSSHHNHNHNPNNTYSPFTTPTQYHPATSTSPSSTAAAAALASPYSSSGHHN
DPSAFSIPQTPPSFDFSANAKWADSVLLEAARAFSDKDTARAQQLWTLNELSSPYGDTE
QKLASYFLQALFNRMTGSGERCYRTMVTAATEKTCSEFSTRKTVLKFQVSPWATFGHV
AANGAILEAVDGEAKIHIVDISSTFCTQWPTLLEALATRSDDTPhLRLLTVVANKFVND
QTASHRMKEIGNRMEKFARLMGVPFKNIIHHVGDLEFDLNELDVKPDEVLAINECVGA
MHGIASRGSPRDAVISSFRRLRPRIVTVVEEADLVGEEGGFDDEFLRGFGECLRWFRV
CFESWEESEFPRTSNERLMLERAAGRAIVDLVACEPSDSTERRETARKWSRMRNSGFGAV
GYSDEVADDVRALLRRYKEGVWSMVQCPDAAGIFLCWRDQPVVWASAWRPT*

>G1778 (1..627)

ATGATGGGATACCAAACAACTCTAATTTCTCCATGTTTTTTTCTCGAAAATGACGAC
CAAAACCACCACAACACTACGATCCTTATAATAATTTCTCTCATCAACTTCTGTTGATTGC
ACTCTCTCACTTGGAAACACCCTCTACTCGTCTCGACGACCACCATAGATTTTCTTCTGCT
AATTCTAACACATCTCCGGCGACTTTTATATTACGGAGGAAACGCTAAGACTTCTTCG
TACAAGAAGGGTGGTGTGCTCATAGCCTACCTCGCCGTTGTGCTAGCTGCGACACCACT
TCAACTCCTCTATGGAGAAACGGACCAAAAGGACCTAAGTCGTTATGTAACGCGTGTGGA
ATCCGATTCAAGAAAGAGGAGAGCGGTGCGACGGCCAGAACTTAACGATCTCCGGTGGA
GGTTCATCAGCGGCAGAAAGTCCAGTAGAGAATTCGTACAACGGAGGTGGAACATATTAC
AGTCATCATCATCACTATGCCTCGTCTCGCCGTCGTGGGCTCATCAGAACACACAA
AGAGTTCCATATTTCTACCGGTTCCGGAGATGGAATATCCCTACGTGGATAACGTCACG
GCTTCTTCTTTTATGTCCTTGGAAATGA

>G1778 Amino Acid Sequence (domain in AA coordinates: 94-119)
MMGYQTNSNFSMFFSSSENDDQNHNYDPYNNFSSSTSVDCITSLGTPSTRDDHHRFSSA
NSNNISGDFYIHGNAKTSSYKKGVAHSLPRRCASCDTSTPLWRNGPKGPKSLCNACG
IRFKKEERRATARNLTISGGSSAAEVPVENSYNNGGNYSHHHHHYASSSPSWAHQNTQ
RVPYFSPVPEMEYPYVDNVTASSFMSWN*

>G1789 (108..413)

CAAGGACTCTGCGACATCTGTGCAACATATCATTTCTCAGAATCTCTTTCTTTCTAGG
TTTATTACTACACAAAACCAACATCATCAACTTTAGTTACTAAACAATGGCATCAGGCT
CAATGTCTTCTTATGGCTCTGGCTCATGGACTGTTAAGCAGAACAAAGCCTTTGAGCGTG
CTCTAGCAGTCTATGACCAAGACACTCCGGACCGTTGGCACAATGTTGCTAGAGCTGTTG
GTGGTAAACACCAAGAGAAGCTAAGAGACAGTATGACCTTCTAGTTCTGTGACATCGAAA
GCATCGAGAAATGGTCACGTGCCATTCCTGACTACAAGACTACTACAGGAAACAGCAACA
GAGGCAGGCTGCGTGATGAGGAAAAGAGGATGAGAAGCATGAAGCTGCAGTGAGACAAGA
AGCAACAAAACCTAACTACGTATGATCGTCAAAATAAAGAGAATCACTTCAGAGAGATG
TGTTTTTTTCAATGTCTGACGAATCAATGTTTTTTTCTTGCAATTTCTCATGTTTTTCCC
TAAGAAATGGTTTTTTTTTCGAGGCAACAAAAA

>G1789 Amino Acid Sequence (domain in AA coordinates: 1-50)
MASGSMSSYSGSWTVKQNKAFERALAVYDQDTPDRWHNVARAVGGKTPEEAKRQYDLLV
RDIESIENGHVFPDYKTTTGNISNRGLRDEEKRMRSMKLQ*

>G1790 (63..1346)

GAAAAAGACTTCACTTTTTTTTTTACTAATTAATTAGTTTTTTTTTCTCCTTTCCAAAA
CAATGGAGAATTTTCGTGACGAGAATGGTTTGTCTCTTAAACAAAACATCTTCACAC
GTGATCAAGAACACATGAAAGAAGAAGATTTCCATTGAGTTCGTGACCAATCAAAAC
CTACAAGCTTTCTTCAAGATTTTCACCATCTTGATCATGATCATCAGTTTGATCATCATC
ATCATCATGGCTCCTCATCTTCCATCCCTTTGCTCAGCGTCCAAACTACGTCTTCTTGTA
TCAATAATGCTCCTTTTCGAGCATTGCTCTTACCAAGAAAACATGGTCGATTTCTATGAAA
CTAAACCAATTTTGATGAATCATCATCATTTTCCAAGCAGTGGAAAACCTACACTTCACTC
GTAATCATCATCATCATCAAGAGATCAATTTGGTCGATGAACATGATGATCCTATGGACT
TGGAGCAAAACAACATGATGATGATGAGGATGATCCCTTTTGATTACCCTCCTACAGAGA
CTTTCAAACCTATGAACCTTCGTAATGCCAGATGAAATTTTCATGTGTTTCTGCAGATAATG
ATTGTTATAGAGCAACGAGTTTCAACAAGACCAAAACATTTCTTACACGAAAGTTGTCTT
CTTCTTCTTTCATCATCATCATGGAAGAAGACCAAAAGTCAACCTTAGTCAAAGGACAAT
GGACTGCTGAAGAAGACAGGGTACTGATTCAACTCGTGGAGAAGTATGGATTGCGTAAAT
GGTCGCATATCGCTCAAGTGTTTACCGGGAAGAATCGGGAACAATGTAGAGAGAGGTGGC
ATAACCATTGAGACCTGACATTAAGAAAGAAAACATGGAGTGAAGAAGAGGACAGAGTGT
TGATAGAATTTCAAAAGAGATTGGAACAATGGGCAGAGATTGCGAAAAGACTCCCGG
GAAGAACAGAGAACTCGATCAAGAACCATTGGAACGCAACAAAAGACAATTCTCTA
AAAGAAAATGTAGATCTAAGTATCCAAGACCTTCTCTGTTGCAGGATTACATCAAGAGCT
TGAATATGGGAGCTTTGATGGCTTCTTCTGTTCTGCAAGAGGTAGACGCAGAGAGAGTA
ATAACAAGAAGAAGGATGTTGTTGTTGCGGTTGAGGAGAAGAAGAAGGAAGAGAGGTGT
ATGGACAAGACAGGATTGTGCTGAAATGTGTGTTTACTGATGATTTTGGATTCAATGAGA
AGCTGCTTGAGGAAGGATGTAGCATTGACTCTTTGCTTGATGACATTCTCAGCCTGACA
TTGATGCTTTTGTTCATGGGCTCTGATTTGTATTTTTTATTCTGCTTGTTCAGTTTTGT
TGTTTTTTGTTTGTCTTTTTTATACGAGACAGATTCCACCAAACCTCAATAATTTGAAAAG
ATATAAAATATTTTGTCTTTTTTAAAAA

>G1790 Amino Acid Sequence (conserved domain in AA coordinates:217-316)

MENFVDENGFA SLNQNI FTRDQEHMKEEDFPFEVVDQSKPTSFLQDFHHL DHDH QFDHHH
HHGSSSSHPLLSVQTTSSCINNA PFHCSYQENMVDFYETKPNLMNHHHFQAVENSYFTR
NHHHHQEI NLVDEHDDPMDLEQNMMMMRMIPFDYPPTE TFKPMNFVMPDEISCVSADND
CYRATSFNKT KPFLTRKLSSSSSSSWKETKKSTLVKGQWTA EEDRVLIQLVEKYGLRW
SHIAQVLPGRIGKQCRERWHNHLRPDIKKETWSEEDRVLIEFHKEIGNKWAEIAKRLPG
RTENSIKNHWNATKRRQFSKRKCRSKYPRPSLLQDYIKSLNMGALMASSVPARGRRRESN
NKKKDVVVAVEEKKKEEVYQDRIVPECVFTDDFGFNEKLLEEGCSIDSLLDDIPQPD
IDAFVHGL*

>G1791 (36..455)

ATGTACATGCAAAAACAAAACCTTAAAGCTTTTCATGGAACGTATAGAGTCTTATAACA
CGAATGAGATGAAATACAGAGGCGTACGAAAGCGTCCATGGGGAAAATATGCGGCGGAGA
TTTCGCGACTCAGCTAGACACGGTCTCGTGTTTGGCTTGGGACGTTTAAACACAGCGGAAG
ACGCGGCTCGGGCTTATGATAGAGCAGCTTTCGGCATGAGAGGCCAAAGGGCCATTCTCA
ATTTTCTCTACGAGTATCAAATGATGAAGGACGGTCCAAATGGCAGCCACGAGAATGCAG
TGGCTTCTCGTCTCGGATATAGAGGAGGAGTGGTGGTGATGATGGGAGGGAAGTTA
TTGAGTTTCGAGTATTTGGATGATAGTTTATTGGAGGAGCTTTTAGATTATGGTGAGAGAT
CTAACCAGACAATTGTAACGACGCAAAACCGCTAGATCATCACTACTTACTTACAGTGTA
ATGTTTTTGGAGTAAAGAGTAATAATCAATATAATATACTTTAGTTTAGGAAAAA
AAAAA

>G1791 Amino Acid Sequence (domain in AA coordinates: TBD)

MERIESYNTNEMKYRGVRKRPWGKYAAEIRDSARHGARVWLGTFN TAEDAARAYDRAAFG
MRGQRAILNFPHEYQMMKDGPNGSHENAVASSSSGYRGGGGDDGREVIEF EYLDSDLLE
ELLDYGERSNQDNCNDANR*

>G1793 (59..1783)

AGTGATTTATTGATTAACCCAAACAAAAATAAACAGATTTGACTCAAAAAGAAGAAAAT
GAATTCATAACACTGGCTTGGCTTTCCTCTTTCACCGAACAACCTTCTTTGCCTCCTCA
TGAATACAACCTTGGCTTGGTCAGCGACCATATGGACAACCTTTTCAAACACAAGAGTG
GAATATGATCAATCCACACGGTGGAGGAGAGATGAAGGAGGAGAGGTTCCAAAAGTGGC
CGATTTTCTCGGTGTGAGCAAACCGGACGAAAACCAATCCAACCACCTAGTAGCTTACAA
CGACTCAGACTACTACTTCCATACCAATAGCTTGATGCCTAGCGTCCAATCAAACGATGT
CGTTGTAGCAGCTTGTGACTCCAATACTCCTAACAAACAGTAGCTATCATGAGCTTCAAGA

>G1793 Amino Acid Sequence (conserved domain in AA coordinates:179-255, 281-349)

>G1795 (27..422)

>G1795 Amino Acid Sequence (domain in AA coordinates: 12-80)

>G1800 (61..894)

93

MEKSSSMKQWKKGPARGKGGPQNALCQYRGVRQRTWGWVAEIREPKKRARLWLGSFATA
BEAAMAYDEAALKLYGHDAYLNLPHLQRNTRPSLSNSQRFKWVPSRKFISMPFSCGMLNV
NAQPSVHI IQORLEELKKTGLLSQSYSSSSSSSTESKNTSF LDEKTSKGETDNMFEGGQD
KKPEIDLTFELQQLGILKDNEAPEAECVSAECHSPPPWNEQEETGSPFRTENFSDWTLIEM
PSETTTMQFDSSNFGSYDFEDDVSFPSIWDYYGSLD*

ATGCAGAGCAGCTTCAAACCGTTTCCTTCACTCCTGATTCTACTCTCAATCCTCTTAC
TTCTTCAGAGGAGATAGTTGTCTTGAGGAGTTTCATCAACCAGTCAATGGTTTTCACCAT
GAAGAAGCTATCGATTAAAGTCCAAATGTCACATATTGCTTCAGCTAACCTTACACTACAC
ACGTTTGATACGGTTATGGATTGTGGTGGTGGTGGTGGCTTGAGGGAGAGACTTGAA
GGAGGAGAAGAGGAGTGTTTGACACAGGGCAATTAGTGTACCAAGAAAGGACAAGATTA
GTAGGAGGAGGAGTAGGAGAAGTGAACAGCAGATTTGGTGTGATTCGGTTTCAGCTATGGCT
GATAACAGTCAACATATCTGACACTTCCACAGATATTGATATCTGATGACAAGACTCAGTTG
AATGAGGCTCATCAAGGGATGCTATTGGCTACAAATTGTTTCAGATCAATCCAATGTGAAA
TCTAGTGATCAAAGGACACTTCGTCGACTTGCTCAGAACCGGGAGGCTGCTAGGAAAAGT
CGTTTGAGGAAAAAGGCCTATGTTTCAGCAACTTGAGAATAGTCGAATCAGGCTTGCACAG
CTAGAGGAAGAGCTCAAAGAGCTCGCCACAGGGATCTTTGGTTGAAAGAGGAGTTTCA
GCGGATCACACGCATTTGCGCAGCAGGAAATGGTGTCTTTTCATTTGAATTGGAATATACA
CGTTGGGAAGGAGGAACATCAAAGTAATGATCAACGACTTAGATCGGGTGTGAATTGCGAC
TTAGGTGACAACGATCTACCGGTTCTAGTGGCTGTGTGATGAGTCACTATGATGAAATA
TTCAGGCTAAAGGGAATTGCGACTAAAGTTGAAGTCTTTCATATGCTCTCAGGCATGTGG
AAGACACCTGCCGAGAGATTTTTCATGTGGTTAGGTGGATTTAGATCATCAGAGTTACTT
AAGATATTGGGGAACCATGTGGATCCATTGACGGACCAGCAGTTGATAGGCATTGCAAC
CTTCAGCAATCGTCTCAACAAGCAGAGGATGCATTGTCCAAAGGCATGGAAGCTCTACAA
CAATCACTTCTCGAGACGCTTTCTTCTGCTTCTATGGGTCCAAACTCTTCAGCAAATGTT
GCAGATTATATGGGTATATGGCTATGGCTATGGGCAAACTTGGCACTCTTGAAAACCTC
CTTCGCGCAGGCTGATTTATTGAGGCAACAACCTGCAACAGCTTTCAGCAAGTTCTCAACC
ACACGACAGCTGTCTCGCGCTTTTGTGTCATCCACATTTATATTTCTCGGCTTAGAGCA
CTTAGCTCTCTATGGTTAGCCAGACCTAGAGACTAA

MSQSSFKTVFPTPDFYSQSSYFFRGDSCLEEFHPQVNGFHHHEEAIDLSPNVTIASANLHYT
TFDITVMDCGGGGGLRERLEGGEEECLDTGQLVYQKGTRLVGGGVGEVNSSWCDSVSAMA
DNSQHTDSTDIDTDDKTQLNGGHQGMLLATNCSDQSNVKSSDQRTLRLLAQNREARKS
RLRKKAYVQQLENSRIRLAQLEEBELKRARQQGSGLVERGVSADHTHLAAGNGVFSFELEYT
RWKEEHQRMINDLRSGVNSQLGDNDLRVLVDAMVSHYDEIFRLKGIGTKVEVFHMLSGMW
KTPAERFFMWLGGFRSELLKILGNHVDPLMTDQQLIGICNLQSSQQAEDLASQMEALQ
QSLELTSSASMGPNSSANVADYMGHMAAMGKLTLENFLRQADLLRQQLRQTLQHLRIIT
TRQAAARFLVIHDYISRLRALSSLWLARPRD*

AAAGGAGCATTTGGTATCTCAAACAATATTTGCCCTTTCTCTATCTCTCTCATCACTAT
TTGCCATCTCTTTCTCTCTCCCTCTCTTTCAAATGTCAATAAACCAATACTCAAGCGATT
TCCACTACCAATTCTCTCATGTGGCAACAACAGCAGCAACAACAACAACACCAAAACGACG
TCGTGGAAGAAAAGAAGCTCTTTTGAGAAAACCTTAACCCCAAGTGACGTCGGAANAAC
TCAACCGCCTCGTCATCCCAAAACAGCAGCCGAGAGATATCTTCCCTAGCTAGCGCCGCCG
CCGACAGCGCCGTGGAGAAAGGACTTCTCCTCTGCTTTGAGGACGAGGAAGGTAAACCAT

GGAGATTCAAGTACTCGTACTGGAACAGTAGCCAGAGTTATGTCTTGACCAAAGGCTGGA
GCAGATACGTCAAGGAGAAGCACCTTGACGCCGAGACGTCGTTCTCTTCCATCGACACC
GTTCAAGACGGCGGAAGATTCTTCATTGGCTGGAGAAGACGCGGTGACTCTTCTTCTCCT
CCGACTCTTATCGCCATGTTCAATCCAATGCCTCGCTCCAATATTATCCTCATGCAGGGG
CTCAAGCGGTGGAGAGCCAAAGAGGCAACTCGAAGACATTAAGACTGTTTCGGAGTGAACA
TGGAGTGCCAGCTAGATTTCGGACTGGTCCGAGCCATCCACACCTGACGGTTCTAACACAT
ATACAACCAATCACGACCAGTTTCTATTTCTACCTCAACAACAACACTATCCTCCTCCGT
ACTACATGGACATAAGTTTACAGGAGATATGAACCGACGAGCTAGAAGCCCACAAGGA
TTAAAAAAAGCTTCACATCTGGTCTGTTATGTTGTCATAGATGTTGATTCTTCTTAATTT
TACACAAGCTTCAATTTTGCATTATTTAAAGTAAATCGTATTTTGATTCTTCTTTAAATC
TCTCTCAATTTTCACTCTCTTCTTTCTTCTTATGTATTAGATTCTTTTACATAGCTA
ACACTTGTATAGAGAATTCAAAGTTCTGGCTATTTTCGAAAGTTATCTTTTCTCTTAAAA
AAAAAAA

>G1811 Amino Acid Sequence (domain in AA coordinates: TBD)

MSINQYSSDFHYHSLMWQQQQQQQHNDVVEEKEALFEKPLTPSDVGKLNRLVIPKQHA
ERYFPLAAAAADAVEKGLLLCFEDEEGKPWRFRYSYWNSSQSYVLTKGWSRYVKEKHLDA
GDVVLPHRRSDGGRFFIGWRRRGDSSSSSDSYRHVQSNASLQYYPHAGAQAQAVESQRGNS
KTLRLFGVNMECQLDSDWSEBSPDPGDSNTYTTNHDQFHFYPQQQHYPYPPYMDISFTGDM
NRTS*

>G182 (74..1366)

CGTCGACGATCAGATTCTTGCGTATAGCTGTATATATACACCAAGATACACTCATCATCG
TCATATATAGATTATGTGCGAGCTCTCTGAGCTTCTTGACATGGAAAACCTTCCAAGGAGA
CTTAACCGACGTCGTACGAGGAATCGGAGGCCACGTGTTATCACCGGAGACTCCTCCCTC
GAACATCTGGCCTCTTCTCTGTACATCCAACACCATCACCGTCAGATCTTAACATAAA
CCCCCTTCGGAGATCCCTTTGTGAGCATGGACGATCCACTCCTCCAAGAACTAAACTCCAT
CACAACTCCGGCTATTTCTCCACCGTAGGAGATAACAACAACAACATTACAACAACAA
TGGTTTCTTGGTTCCAAAGGTATTTGAGGAGGATCATATAAAGAGTCAATGTAGTATCTT
CCCAAGAAATCCGGATCTCGCATAGTAACATCATCCAGATTCTTCTCCGTGTAATTCTCC
GGCATGTGCGGCTCAGTGTGTCGAGCCGAGCAGCCGCTCGCCGAGAGGCATCATCAA
CGTAGACACAAACAGTCTCTAGAACTGTCTATTGGTTGATGGTACCAGTTCTCCTCGCA
GATTCAATATCTTCCCTCGGAATCTAGGCCTTAAAAGAAGGAAGAGTCAGGCAAAGAA
GGTGGTGTGATTCCGGCCCCGGCTGCAATGAACAGCCGATCAAGCGGAGAAGTGGTTCC
ATCGGATCTATGGGCTTGGCGTAAATACGGTCAAAAACCTATCAAAGGCTCTCCTTTTCC
AAGGGGTTATTATAGATGCAGCAGCTCAAAAGGTTGTTCAAGAAAGCAAGTCGAAAG
AAGCCGAACCGATCCAACATGTTGGTGATTACATATACCTCCGAACATAACCATCCTTG
GCCCATCCAACGCAACGCTCTCGCCGGCTCCACACGCTCCTCCACCTCCTCCTCATCTAA
CCCTAATCCTTCCAACCCCTCAACCGCAAACGTAACCTCCTCATCCATTGGCTCCCAAAA
CACCATCTACTTGCCTTCTCCACCACTCCTCCTCTACCCTCTCATCTCCGCCATCAA
AGATGAACGAGGGGACGATATGGAGTTGGAAAACGTAGATGATGATGATAACCAGAT
TGCTCCATACAGACCGGAGCTTCATGATCATCAGACCAACCAGATGATTCTTTGCAGA
TCTTGAAGAGCTAGAAGGAGATTCTTAAGCATGTTGCTTTCTCATGGCTGTGGCGGCGA
CGGGAAGGATAAAACGACCGCGTCCGATGGGATCAGCAATTTCTTCGGGTGGTTCGGGAGA
TAATAATTATAAATAATTACGACGACCAAGACTCAAGGTCGTTATAGTATAGTGTTAATTA
CAGGTAAACAAATTATATTAAATTAAGTTGAGCTTGTGAAAATGAAGATCATATGGTCTG
GTCAGGTTGGGGGC

>G182 Amino Acid Sequence (conserved domain in AA coordinates:217-276)

MCSVSELLDMENFQGLDLDVVRGIGGHVLSPETPPSNIWPLPLSHPTSPSDLNINPFGD
PFVSMDDPLLQELNSITNSGYFSTVGDNNNNIHNNGFLVPKVFEEDHIKSQCSIFPRIR
ISHSNIHDSFPCNSPAMSAHVAAAAASPRGIINVDTNsprncllvdgttfssqIQIS
SPRNLGLKRRKSQAKKVVICIPAPAMNSRSSGEVVPDLWAWRKYGQKPIKGSPPFRGY
RCSSSKGCSARKQVERSRTDPNMLVITYTSEHNHPWPIQRNALAGSTRSSTSSSSNPNPS
KPSTANVNSSSIGSQNTIYLPSTTPPTLSSSAIKDERGDDMELENVDDDDNQIAPYR
PELHDHQHQPDDFFADLEELEGDSLMLLSHGCGDGKDKTTASDGISNFFGWSGDMNIN
NYDDQDSRSL*

>G1835 (1..969)

ATGATTGGAACAAGCTTCCCCGAGGATCTTGATTGTGGCAACTTCTTTGACAACATGGAT
GATCTCATGACTTTCCCGGTGGAGATATCGATGTCGGTTTCGGCATAGGTGACTCCGAC

TC'TTTCCTACCATCTGGACCACTCATCACGACACGTGGCCTGCCGCTTCTGATCCTCTC
TTCTCTTCCAACACCAACTCTGATTTCATCACCTGAGCTCTATGTTCCGTTTGAGGACATT
GTTAAGGTGGAAAGACCTCCAAGCTTTGTAGAGGAAACATTGGTTGAGAAGAAGGAAGAT
TCGTTTTTCGACAAACACTGATTTCATCATCTTCTCATAGCCAATTTCAGGAGCTCAAGTCCA
GTGTCGGTTCTCGAGAGCAGCTCCTCCTCGTCTCAAACCACCAACACAACCTCCCTTGTT
CTCCCTGGAAAGCACGGTCGTCCACGCACAAAACGCCCTCGTCCACCTGTCCAGGATAAA
GATAGAGTCAAAGACAATGTGTGCGGTGGTGA CTGCGCCTCATCATTAGAATACCGAAA
CAGTTTCTCTCTGATCACAACAAGATGATCAACAAGAAGAAGAAGAAGAGGCCAAGATT
ACTTCTTCTCTTCTTCTCGTCCGGGATTGATCTTGAAGTCAATGGAAACACGTCGATTCTG
TATTCTTCAGAGCAATATCCGCTTAGGAAATGTATGCACTGTGAGGTACCAAGACTCCA
CAGTGGAGGCTTGGTCCAATGGGTCCAAAGACACTTTGCAATGCGTGGGTGTACGTTAC
AAATCAGGGAGGCTTTTCCCGGAGTACCGTCCAGCTGCTAGTCCAACATTTACTCCAGCT
CTTCACTCAAACACACAAGAAAGTGGCTGAAATGAGAACAAGAGATGCAGTGATGGT
AGCTACATAACCGAAGAGAATGATCTGCAAGGGCTGATTCCGAACAATGCCTACATTGGC
GTAGACTAA

>G1835 Amino Acid Sequence (domain in AA coordinates: 224-296)
MIGTSFPEDLDCGNFFDNMDDLMDPFGDIDVGFIGDSDFSFTIWTTHHDTWPAASDPL
FSSNTNSDSSPELVVPFEDIVKVERPPSFVEETLVEKKEDSFSTNTDSSSSHSQFRSSSP
VSVLESSSSSSQTNTTSLVLPKGHRPRTKRPRPPVQDKDRVKNVCGGDSRLIIRIPK
QFLSDHNKMKKKKKKAKITSSSSSSGIDLEVNGNNVDSYSSEQYPLRCKMHCEVTKTP
QWRLGPMGPKTLNACGVRYKSGRLFPFYRPAASPTFTPALHSNSHKKVAEMRNKRCSDG
SYITEENDLQGLIPNNAYIGVD*

>G1836 (47..610)
ATAACAAGCCTAGAACACTAGAACTTCAAAAAAGAAAAAATCTTATGGAGAACAACAA
CGGCAACAACCAGCTGCCACCGAAAGGTAACGAGCAACTGAAGAGTTTCTGGTCAAAAGA
GATGGAAGGTAACCTTAGATTTCAAAAATCACGACCTTCCTATAACTCGTATCAAGAAGAT
TATGAAGTATGATCCGATGTGACTATGATAGCTAGTGAGGCTCCAATCCTCCTCTCGAA
AGCATGTGAGATGTTTATCATGGATCTCACGATGCGTTCGTGGCTCCATGCTCAGGAAAG
CAAACGAGTACAGCTACAGAAATCTAATGTGCGATGCCGAGTGGCTCAAACCTGTTATCTT
TGATTTCTTGCTTGATGATGACATTGAGGTAAAGAGAGAGTCTGTTGCCGCCGCTGCTGA
TCCTGTGGCCATGCCACCTATTGACGATGGAGAGCTGCCTCCAGGAATGGTAATTGGAAC
TCCTGTTTGTGTAGTCTTGAATCCACCAACCACAACCAAAATGCAGGCATGGCCTGG
AGCTTGGACCTCGGTGTCTGGTGAGGAGGAAGAAGCGCGTGGGAAAAAAGGAGGTGACGA
CGGAAACTAATAAGTGAATACGTTTTAGGGTATTTTCAAGGGAATATGTAGTAAATAGT
CATGGATC

>G1836 Amino Acid Sequence (domain in AA coordinates: 30-164)
MENNNNGNNQLPPKNEQLKSFWSKEMEGNLDPFNHDLPITRIKKIMKYDPDVTMIASEAP
ILLKACEMFIMDLTMRWLHAQESKRVTLQKSNVDAAVAQTVIFDFLLDDDI EVKRESV
AAAADPVMPPIDDELPPGMVIGTPVCCSLGIHQPPQMQAWPGAWTSVSGEEEEARGK
KGGDDGN*

>G1838 (132..1628)
TTCTTGGCATTCTCTTTAGAACTTTTCGTACAAAATGCAAAACCTGAACCTCTAAAGCTA
AAAAAAGAGATTAGAGACTGTAAGTCTTTTATCAGATTTTCAACTAGGAAAAAAGTTAC
AATCTTTTTTGATGGCTCCTCCAATGACGAATTGCTTAACGTTTTCTCTGTACCAATGG
AGATGTTGAAATCAACTGATCAGTCTCACTTCTCTTCTTCTTACGACGATTCTTCTACTC
CTTATCTCATCGATAACTTCTATGCTTTCAAAGAAGAAGCTGAGATAGAAGCTGCTGCTG
CTTCAATGGCGGATTCAACAACCTTATCTACTTTTTTCGATCATTCTCAGACTCAGATTCT
CAAAGCTGGAAGATTTCTCGGTGATTCTTTGTCCGTTACTCTGATAACCAACAGAGA
CCCAAGACTCTTCTCTCTCACTCCATTCTACGATCCACGTCACCGCACCGTTGCCGAAG
GAGTTACAGGGTTCTTCTCTGATCATCATCAGCCAGATTTCAAGACGATAAACTCGGGAC
CAGAAATCTTCGATGACTCAACAACCTTCCAACATCGGTGGTACTCATCTCTCCAGTCACG
TGGTGGAGTCATCAACGACGGCGAAGTTAGGGTTAACGGTGATTGCACCACCACCGGAG
GAGTTTTGTCTCTAGGGGTTAACAACACATCAGATCAACCTTTGAGCTGTAACAATGGCG
AGAGAGGTGGAACAGTAACAAGAAGAAACAGTTTCTAAGAAGGAACATCAGATGATT
CAAAGAAGAGATTGTGCAACATTGGGACAAAGAAGTTCAATTTATCGTGAGTCAACC
GACATAGATGGACTGGAAGATACGAAGCGCATCTATGGGATAACAGCTGTAGGAGGGAAG
GTCAAGCCAGAAAAGGACGTCAAGTGTACTTAGGTGGATATGACAAGGAAGATAGAGCAG

CTAGAGCCTATGACTTGGCAGCTTTAAATACTGGGGTTCTACTGCTACTACAAATTTTC
CGGTCTCGAGTTATTCAAAAGAACTTGAGGAAATGAATCACATGACCAAGCAAGAGTTTA
TTGCATCTCTTAGGAGGAAAAGTAGCGGTTTTTCGAGAGGAGCTTCAATATATAGAGGTG
TCACAAGGCATCATCAACAAGGTCGCTGGCAAGCAAGAATCGGCCGTGTCGCAGGAAACA
AAGATCTTTACCTCGGAACCTTTTGCAACCGAAGAGGAAGCAGCAGAGGCTTATGACATTG
CAGCCATAAAGTTTCAAGGAATCAACGCAGTAACTAATTTGAGATGAACAGGTATGACA
TTGAAGCTGTATGAATAGTTCTTTACCTGTAGGAGGAGCAGCTGCGAAACGCCACAAAC
TCAAACTCGCTCTTGAATCTCCTTCTTCATCATCCTCTGACCATAACCTCCAACAACAAC
AGTTGCTTCCGTCTCTTCTCCCTCGGATCAAAACCTAATCAATCCCATGTGGCATTCTC
CATTTGAGCCTTCAGTTCTCTATTACCACCAGAACTTCTTTCAGCATTTATCCTTTGGTCT
CTGACTCTACAATTCAAGCTCCTATGAACCAAGCTGAGTTTTTCTTGTGGCCTAACAGT
CTTACTAAATCATTTGGTTCTTCTTGTCTTAGACTTCTATTACCGCACTAACCGATGAC
CCGAGGCTTATCTTTGATTCTGGCTATAAGGATGAATCTTTCAAGTTCTTTTTTAAC
TGTAGGCTAAGACAGAAGTAGAGGGGAGAAAAGTTGAAGAATCTGAACTTTTGGGGTCA
ATTTTGTATTAATGTTTTTCTTTTGTCAAGGGTGGATTATCGGTTTTATTACTTATTTTT
TGAATGTAATCGGCCTATAACGGTATAACTCTGTTTCCATTTATGAATATTTTCTCAA
TTGAAAAAAAAAAAAAAAAAAAA

>G1838 Amino Acid Sequence (conserved domain in AA coordinates: 229-305, 330-400)
MAPPMNTNCLTFSLSPEMLKSTDQSHFSSSYDDSTPYLIDNFYAFKEEAEIEAAAASMA
DSTTLSTFFDHSQTQIPKLEDFLGDSFVRYSDNQETQDSSSLTPFYDPRHRTVAEGVTG
FFSDHHQDPFKTINSQPEIFDDSTTSNIGGTHLSSHVESSTAKLGFNGDCTTTGGVLS
LGVNNTSDQPLSCNNGERGGNSNKKKTVSKKETSDDSKKIVETLGQRTSIYRGVTRHRW
TGRYEHLWDNSCRREGQARKGRQVYLGGYDKEDRAARAYDLAALKYWGSTATNFPVSS
YSKELEEMNHMTKQEFIASLRKSSGFSRGASIYRGVTRHHQQGRWQARIGRVAGNKDLY
LGTFAEEEEAEAYDIAAIKFRGINAVTNFEMNRYDIEAVMNSSLPVGGAAAKRHKLKLA
LESPSSSSSDHNLQQQLLPSSSPSDQNPNSIPCGIPFEPVSVLYYHQNFFQHYPLVSDST
IQAPMNQAEFFLWPNQSY*

>G1843 (51..653)
CAGACATCACAATCAAAATTAGGTCAGAAGAATTAGTCGGAGAAAACAGCCATGGGAAGAA
GAAAAGTAGAGATCAAACGAATTGAGAACAAAAGCTCTCGACAAGTTACTTTCTGTAAAC
GACGAAATGGTCTCATGGAGAAAGCTCGTCAACTCTCAATTCTTTGTGAATCTCCGTCG
CTCTTATCATCATCTCTGCCACCGAAGACTCTACAGCTTCTCCTCAGGTGATAGCATGG
CCAAGATCCTCAGTCGTTATGAATTAGAACAGGCTGATGATCTTAAACCTTGGATCTAG
AAGAAAAAATCTTAATTATCTTTCGCACAAGGAGTTGCTAGAAACAATCCAATGCAAGA
TTGAAGAAGCGAAAAGCGATAATGTAAGTATAGATTGTCTAAAGTCCCTGGAAGAGCAGC
TCAAGACTGTCTGTCTGTAAGTCTAGAGCTAGGAAGACAGAACTAATGATGGAGCTTGTGA
AGACCCATCAAGAGAAGGAGAAGCTGCTGAGAGAGGAGAACCAGAGTTTGACTAACCAGC
TTATAAAGATGGGGAAGATGAAGAAGTCTGTGGAAGCAGAGGATGCAAGAGCAATGTCAC
CGGAAAGTAGCTCTGACAACAAGCCACCGGAGACTCTCCTGCTTCTCAAGTAACCACCAT
CACCAACGACTGATTCGAAAAATAAAAATTGTAAAAATTATGATTTGTAGTTCATAAGGA
AAGCTACATACTGTATGTTAAAAATCCTCTTCTCCCTGCTACGGAAAAGTCATCCAA
GGAGATGCATCAATAAAGTAATTGATTTTTATTGTTA

>G1843 Amino Acid Sequence (domain in AA coordinates: 2-57)
MGRRKVEIKRIENKSSRQVTFCKRRNGLMEKARQLSILCESSVALIIISATGRLYSFSSG
DSMAKILSRYLEQADDLKTLDLEEKTLNLYLSHKELLETIQCKIEEAKSDNVSIDCLKSL
EEQLKTALSVTRARKTELMELVKTHQEKEKLLREENQSLTNQLIKMGKMKKSVEAEDAR
AMSPSSSDNKPPETLLLLLK*

>G1853 (1..186)
ATGAGAGGTTCTTGGTACAAGAGTGTTTCTCTGTTTTTGGTCTCAGACCACGGATCAGA
GGGTTGTTATTCTTCATTGTTGGTGTGTGGCTCTAGTTACTATTTTAGCACCATTGACA
TCTAATTCGTATGATTCTTCGTCAAGTTCGACACTTGTGCCGAACATTTATAGTAACTAT
AGGAGGATAAAGGAGCAAGCTGCTGTTGATTATCTTGATCTGAGGTCTCTTTCTTTAGGG
GCTAGTTTTAAAGAGTTTTCTTTTTTGTGGTAAAGAAAGAGAAAGTTATGTGCCTTGTAT
AACATAACTGGGAATTTGCTTGTGGCTTCAAGAGGGTGAGGAGTTAGATCGACATTGC
GAGTTTGAAGAGAGAAGGAAAGATGTGTAGTTCGTCCTCCGAGAGATTATAAAATACCA
CTTAGGTGGCCACTTGGTAGAGATATCATATGGAGTGGGAACGTGAAGATTACCAAGAC
CAGTTTCTTTCTCAGGAAGTGTGACAACGAGGTTAATGTTGCTTGAAGAGAATCAAATA

ACCTTTCCTACTCGGAGGACGGCCTGGTCTTTGATGGGGTCAAAGACTATGCTCGTCAAATT
GCTGAGATGATAGGTTTAGGAAGTGATACTGAATTTGCTCAAGCGGGTGACGGACTGTG
TTAGACATTGGTTGCGGATTTGGTAGCTTTGGTGTCTCAATTTAGTGTCTTTGAAGCTGATG
CCTATATGTATTGCTGAGTATGAGGCAACTGGGAGCCAAGTTCAGTTAGCTCTAGAGAGA
GGCCTTCTCGCAATGATTGGCAATTTCTTTTCAAACAGCTTCCTTATCCAGCACTGTCT
TTTGATATGGTCCATTGTGTCTCAATGTGGCACTACTTGGGATATCAAAGATGCAATGCTA
CTTTTGGAAAGTGGATCGTGTCTGAAACCCGGGGGATACTTTGTTTTAACTTCTCCCACA
AACAAAGCACAGGGAAACTTACCAGATACCAAGAAAACGAGCATCTCAACACGGGTGAAT
GAGTTATCTAAGAAAATCTGTTGGAGTCTAACAGCTCAGCAGGATGAGACGTTTCTTTGG
CAGAAAACCTTCTGATTCAAGTTGCTATTCTTCTCGTTGCAAGCTTCTATACCTCTTTCG
AAAGATGGAGATAGCGTTCCTGATTACCACCCATTGGTTCCATGTATAAGCGGAACCAG
AGTAAACGCTGGATTCTTATACAGAACAGGTCTGTGTTGCAAGGAACAACCTCTGCCGGG
CTTGAAATTCATGGTTTAAAACCGGAAGAATTCTTCGAGGATACACAAATATGGAGATCA
GCTCTGAAAAACTATTGGTCTTGTCTTACACCTCTAATTTTCTCTGACCATCCGAAGAGA
CCCGGTGATGAGGATCCTCTCCCGCCTTTCAACATGATACGCAATGTGATGGACATGCAT
GCTCGTTTTGGGAATTTAAATGCCGCTTTACTCGACGAAGGAAAATCTGCTTGGGTAATG
AACGTCGTCCTCAGTCAATGCACGTAATACTCTTCTATCATACTTGATCGTGGTTTTGCC
GGTGTCTACATGACTGGTGTGAACCATTCCTCGACATATCCTCGAACATATGACATGCTT
CATGCCAATGAATCTCTCACACATCTTAGCTCAGAACGATGCAGCCTAATGGACTTGTTC
TTGGAGATGGACCGGATCTTCCGCTTGAGGGATGGGTTGTTCTAAGCGACAAAGTGGGA
GTAATCGAGATGGCTCGAGCACTTGCAGCTCGAGTGCCTTGGGAAGCAAGAGTCATTGAT
CTTCAAGATGGTAGTGACCAAGACTTCTCGTCTGTCAAAAACCATTCATCAAAAAATAA
>G1853 Amino Acid Sequence (domain in AA coordinates: entire protein)
MRGSWYKSVSSVFLRPRIRGLLFFIVGVVALVTILAPLTSNSYDSSSSSTLVPNIYSNY
RRIKEQAAVDYLDLRLSLGASLKEFFFCGKERESYVPCYNITGNLLAGLQEGEELDRHC
EFEREKERCVRPPRDYKIPLRWPLGRDIWISGNVKITKDQFLSSGTVTTRLMLLEENQI
TFHSEDLVFDGVKDYARQIAEMIGLGSDEFAQAGVRTVLDIGCGFSFGAHLVSLKLM
PICIAEYEATGSQVQLALERGLPAMIGNFFSKQLPYPALSFDMVHCAQCCTTWDIKDAML
LLEVDRVLKPGGYFVLTSPTNKAQGNLPDTKKTSISTRVNELSKKICWSLTAQQDETFLW
QKTSDDSSCYSSRSQASIPCKDGDSPYYHPLVPCISGTTSKRWISIQNRSAVAGTTSAG
LEIHGLKPEEFFEDTQIWRSAKKNYWSLLTPLIFSDHPKRPGEDEPLPPFNMIRNVMDMH
ARFGNLNAALLDEGKSAWVMNVVPVNARNTLPIILDGRFAGVLHDWCEPFPPTYPRTYDML
HANELLTHLSSERCSLMDLFLEMDRILRPBGWVVLSDKVGVIEMARALAAARVRWEARVID
LQDGSQRLLVLCQKPFIKK*
>G1855 (1..1902)
ATGGCGAAAGAGAACAGTGGTCAATCATCACCAAAACAGAAGCAAGAAGAAAGAACTAACT
TTGATTCTTGGTGTAAGTGGACTCTGCATTTTGTCTATGTTTTAGGTGCATGGCAAGCC
AATACCGTCCCATCTTCTATCTCGAAGCTCGGATGCGAGACGCAATCAAACCTTCTTCG
TCCTCTTCTCTTCTCCTCATCTTCAGAGTCAGCTGAAGTAGATTTCAAAGCCATAATCAG
ATTGAGTTAAAGGAAACAAACCAACCATTAAAGTACTTTGAACCATGTGAATTATCTCTC
AGTGAGTACACTCCTTGTGAAGACCGACAAAGAGGAAGAAGATTGATAGGAACATGATG
AAATATAGAGAAAGACATTGTCTGTAAAAGATGAGCTTCTTTATTGTTTGATTCTCTCT
CCACCAAACTACAAGATTCCATTTAAATGGCCACAAAGTAGAGACTATGCTTGGTATGAC
AATATCCCTCACAAGGAACCTAGTGTGAGAAAGCAGTTCAAACCTGGATTCAAGTTGAA
GGTGACCGCTTTAGATTCCCTGGTGGTGGTACTATGTTTCTCGTGGAGCTGATGCTTAT
ATCGATGATATTGCTAGGCTTATTCCTCTTACTGATGGTGGAAATCAGAACAGCTATTGAC
ACTGGATGTGGTGTGCAAGTTTTGGTGTCTTACCTCTTGAAGAGAGACATTATGGCTGTG
TCTTTTGTCTCAAGAGACACTCATGAAGCTCAGGTACAGTTTGTCTTTAGAACGCGGAGTT
CCTGCGATAATCGGGATTATGGGATCAAGAAGACTTCTTATCCAGCTAGAGCTTTTGAT
CTTGCTCATTGTTCTCGTTGTTTGTATCCCTTGGTTTAAAAATGATGGTTTGATCTTATG
GAGGTCGACCGGGTTTTAAGACCGGGCGGTACTGGATCCTCTCGGGACCACCGATTAACT
TGGAAACAGTACTGGAGAGGGTGGGAGAGAACAGAGGAGGATTTGAAGAAAGAGCAAGAT
TCAATAGAAGATGCAAGAGTCTTTGCTGGAAGAAAGTAACGAAAAAGGTGACTTA
TCAATTTGGCAAAAGCCTCTCAATCACATTGAGTGTAAAAAGCTCAAACAAAAACAATAAG
TCACCTCCGATATGCAGCTCAGATAACGCGGATTCCGCTTGGTACAAAGACTTGGAAACT
TGTATAACACCATTACCAGAAACAAACAATCCAGATGATTGAGCAGGCGGTGCACTCGAG
GATTGGCCAGACCGAGCATTCGCGGTACCTCCAAGAATCATCAGAGGAACATATACCAGAA

ATGAACGCGGAGAAATTTAGAGAAGACAACGAGGTTTGGAAAGAGAGAATAGCACATTAC
AAGAAGATAGTCCCTGAGCTTTTCATGGAAGATTCAGGAACATTATGGACATGAACGCT
TTTCTCGGCGGATTTCGCTGCTTCCATGCTGAAATATCCCTCATGGGTCATGAACGTTGTC
CCGGTCGATGCAGAGAAACAAACGTTAGGTGTGATCTACGAACGTGGATTGATAGGGACG
TATCAAGATTGGTGTGAAGGATTCTCAACGTATCCAAGAACTTATGATATGATTCATGCA
GGAGGATTGTTTCAAGTTATACGAACATAGGTGTGATTGACGTTGATATTGTTGGAGATG
GATCGAATTTTGGAGACCAGAAGGAACAGTTGTGTTGAGAGATAATGTGGAGACGTTGAAT
AAGGTAGAGAAGATAGTGAAGGGAATGAAGTGAAGAGTCAAATTGTTGATCATGAGAAA
GGTCCTTTTAATCCTGAGAAGATTCTGTTGCTGTTAAACTTATTGGACTGGTCAACCT
TCTGACAAGAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA
>G1855 Amino Acid Sequence (domain in AA coordinates: entire protein)
MAKENSQHGHQTEARRKKLTLLGVSLCILFYVLGAWQANTVPSSISKLGCEQSNPSS
SSSSSSSSSESALDFKSHNQIELKETNQTIKYFEPCELSLEYTPCEDRQRGRFRDRNMM
KYRERHCPVKDELCLIPPPNYKIPFKWPQSRDYAWYDNIPHKELSVEKAVQNWIQVE
GDRFRFPGGGTMFPRGADAYIDDIARLIPLTDGGIRTAIDTGCVASFGAYLLKRDIMAV
SFAPRDTHEAQVQFALERGVPALIGIMSSRLPYPARAFDLAHC SRLIPWFKNDGLYLM
EVDRVLRPGGYWILSGPPINWKQYWRGWERTEDLKKEQDSIEDVAKSLCWKKVTEKGD
SIWQKPLNHIECKKLQMNKSPICSSDNADSAWKDLETCITPLPETNNPDDSSAGGALE
DWPDRFAVPPRIIRGTIPEMNAEKFREDEWVKERIAHYKKIVPELSHGRFRNIMDMNA
FLGGFAASMLKYPWVMNVVPVDAEQKTLGVIYERGLIGTYQDWCEGFSTYPRTYDMIHA
GGLFSLYEHRCDLTLILLEMDRILRPEGTVVLRDNVETLNKVEKIVKGMKWSQIVDHEK
GPFNPEKILVAVKTYWTGQPSDKNNNNNNNNNN*
>G187 (118..1074)
TAGACCTCTTAGGAAAAAACCTAAAAACCTAATCCCCAACCTAAAAGGCTTATCTCAT
CTCTTCTTCTTTGTCTTCTTTACTCTTTTACCTCTCTCTTCTATGTTCTTACCATG
TCTAATGAAACCAGAGATCTCTACAACCTACCAATACCCCTTCATCGTTTTCGTTGCACGAA
ATGATGAATCTGCCTACTTCAAATCCATCTTCTTATGGAACCTCCCATCAGAAACGGT
TTTAATCCATCTACTTATCTCTTACCGATTGTCTCAAAGTTCTCCAGCAGCGTATGAA
TCTCTACTTCAGAAAACCTTTGGTCTTTCTCCCTCTTCTCAGAGGTTTTCAATTCTTCG
ATCGATCAAGAACCAGAACCGGTGATGTTACTAATGACGTAATCAATGGTGGTGCATGCAAC
GAGACTGAAACTAGGGTTTCTCCTTCTAATTTCTCCTCTAGTGAGGCTGATCACCCCGGT
GAAGATTCGGGTAAGAGCCGAGGAAACGAGAGTTAGTCGGTGAAGAAGATCAAATTTCC
AAAAAAGTTGGGAAAACGAAAAAGACTGAGGTGAAGAAACAAAGAGACCCAGAGTCTCG
TTTATGACTAAAAGTGAAGTTGATCATCTTGAAGATGGTTATAGATGGAGAAAATACGGC
CAAAAGGCTGTAAAAAATAGCCCTTATCCAAGGAGTTACTATAGATGTACAACACAAAAG
TGCAACGTGAAGAAACGAGTGGAGAGATCGTTCCAAGATCCAACGGTTGTGATTACAACT
TACGAGGGTCAACACAACCACCGATTCCGACTAATCTTCGAGGAAGTTCTGCCGCGGCT
GCTATGTTCTCCGACAGCTCATGACTCCAAGAAGCTTTGCACATGATATGTTTAGGACG
GCAGCTTATACTAAGCGCGGTTCTGTGGCGGCGGCTTTGGATTATGGATATGGACAAAGT
GGTTATGGTAGTGTGAATCAAACCTAGTTCTCACCAGTGTATCATCAAGGGGGTGAG
TATGAGCTCTTGAGGGAGATTTTCTTCAATTTTCTTTAAGCAAGAGCCTTGATCGATC
ATTGTTATACTACATATATATATATATATTGAGAGAGAGAGGTAGAGAAAAAAA
>G187 Amino Acid Sequence (domain in AA coordinates: 172-228)
MSNETRDLYNQYPSFSLHEMMNLPTSNPSSYGNLPSQNGFNPSTYSFTDCLQSSPAAY
ESLLQKTFGLSPSSSEVFNSSIDQEPNRDVTNDVINGGACNETETRVSPSNSSSEADHP
GEDSGKSRKRRELVGEEDQISKVKGKTKKTEVKKQREPRVSFMTKSEVDHLEDGYRWRKY
GQKAVKNSPYPRSYRCTTKCNVKKRVERSFQDPTVVITTYEQHNHPIPTNLRGSSAA
AAMFSADLMTPRSFAHDMFRTAAYTNGGSVAAALDYGQSGYGSVNSNPSSHQVYHQGG
EYELLREIFPSIFFKQEP*
>G1881 (1..519)
ATGCGAATTTTGTGTGATGCTTGTGAGAGCGCCGCGCTATCGTCTTTTGCGCCGCCGAC
GAAGCTGCCCTCTGTTGCTCCTGCGACGAAAAAGTTCATAAGTGCAACAAGCTGGCTAGT
CGGCATCTTCGTGTAGGCTTAGCTGATCCGAGTAATGCACCAAGCTGTGACATATGCGAA
AATGCACCCGCAATCTTTTACTGTGAGATAGATGGTAGTTCCCTTTGTCTACAATGTGAT
ATGGTGGTACATGTTTGGTGGGAAGAGAACACATAGGCGGTTTCTATTACTGAGACAGAGA
ATTGAGTTTCCAGGCGATAAGCCTAATCATGCTGACCAACTGGGACTACGGTGTCAAAG
GCTTCCTCTGGTCGTGGTCAAGAATCAAATGGGAATGGTGTATCATGATCATAATATGATC

GATCTTAACCTCCAATCCTCAAAGAGTACACGAGCCTGGATCACATAACCAAGAGGAGGGT
ATTGATGTAAATAACGCAAACATCACGAGCATGAATAG
>G1881 Amino Acid Sequence (domain in AA coordinates:5-28, 56-79)
MRILCDACESAAAIVFCAADEAALCCSCDEKVKCNKLASRHLRVGLADPSNAPSCDICE
NAPAFYCEIDGSSSLCLQCDMVHVGGKRTHRRFLLLRQRIEFGDKPNHADQLGLRCQK
ASSGRGQESNGNDHDHNMIDLNSNPQRVHEPGSHNQEEGIDVNNANNHEHE*
>G1882 (1..1200)
ATGGTTTTTTTCTTCATTTCTTACTTATCCTGATCATTCATCAAACCTGGCAACAACAACAT
CAACCAATCACAACCACCGTTGGATTACCGGAAATAACATCAACCAACAGTTTCTTCCT
CACCATCCCCTCCCACCGCAACAGCAACAAACGCCCTCCGCAGCTTCACCACAACAACGGT
AACGGCGGAGTCGCTGTTCCCGGTGGACCTGGCGGGTTAATCCGACCAGGTTTCGATGGCG
GAAAGAGCAAGGCTAGCCAACATACCATTACCTGAAACAGCCTTGAAGTGTCCAAGATGT
GACTCAACTAACACCAAATTCGTTACTTCAACAACCTACAGTCTCACTCAACCTCGCCAC
TTCTGCAAAAGCATGCCGTCGTTACTGGACACGTGGCGGTGCTCTAAGGAGCGTTCCCGTC
GGTGGCGGTTGCCGTAGAAACAAAAGAACAAAAACAGCAGCGGTGGAGGTGGCGGTAGC
ACCAGTAGCGGTAACAGCAAGTCACAAGACAGCGCCACGAGCAACGACCAATACCACCAC
CGAGCCATGGCTAACAATCAGATGGGACCACCTTCTTCGTCATCGTCTCTAAGCTCGTTG
CTGTCTTCTTACAACGCAGGGTTAATCCCCGGACATGATCATAACAGCAATAACAACAAC
ATACTTGGACTTGGATCATCTTTGCCTCCTCTTAAGCTTATGCCTCCTTTAGACTTCACA
GACAACTTCACCTTACAATACGGTGCCGTTTCAGCTCCTTCTTATCATATAGGCGGTGGA
AGCAGTGGAGGAGCGCGGCTCTTTTAAACGGTTTTTGACCAGTGGAGATTCCCGGCAACA
AACCAACTTCTTTAGGCGGTTTAGACCCGTTTGATCAACAACATCAAATGGAGCAGCAG
AATCCAGGTTACGGATTGGTTACCGGTCGGGTGAGTATCGACCTAAGAACATTTTCCAT
AACCTTATCTCTCTTCTTCTGCTGCTTCATCAGCTATGGTTACAGCCACCGCTCGCAA
TTAGCTTCAGTGAAGATGGAAGATAGTAACAATCAGCTCAACTTGTCTAGACAACTTTTT
GGAGACGAACAACAGCTCTGGAATATTCATGGCGCTGCTGCAGCATCCACCGCAGCTGCA
ACAAGTTTCGTGGAGTGAAGTCTCTAATAATTTAGTTCTTCTTCTACTAGCAATATATAA
>G1882 Amino Acid Sequence (domain in AA coordinates:97-125)
MVFSSFTPTYPDHSSNWQQHQPIITTVGFTGNNINQQFLPHHPLPPQQQTTPQLHHNNG
NGGVAVPGGPGGLIRPGSMAERARLANIPLPETALKCPRCDSTNTKFCYFNYSILTQPRH
FCKACRRYWTGRGALRSPVVGCGCRNRKTKNSSGGGGSTSSGNSKSDSATSNDQYHH
RAMANNQMGPPSSSSSLSSLLSSYNAGLIPGHDHNSNNNNILGLGSSLPPLKLMPLDFT
DNFTLQYGAVSAPSYHIGGSSGAAALLNGFDQWRFPATNQLPLGGLDPFDQQHQQMEQQ
NPGYGLVTGSGQYRPNKIFHNLISSSSASSAMVTATASQLASVKMEDSNNQLNLSRQLF
GDEQQLWNHIGAAAATAAATSSWSEVSNNFSSSSTSNI*
>G1883 (1..1110)
ATGGACGCTACGAAGTGGACACAGGGTTTTCAAGAAATGATGAACGTTAAACCAATGGAG
CAGATCATGATTCTTAATAACAACACATCAACCAACACCACATCCAATGCAAGGCCA
AACACCATTCTCACATCTAACGGCGTCTCAACTGCTGGAGCAACCGTCTCCGGCGTAAGC
AACAAACAATAACAATACGGCGGTTGTGGCGGAGAGGAAAGCAAGACCACAAGAGAACTA
AATTGTCCAAGATGCAACTCAACCAACACAAAGTTTTGTTACTACAACAATATAGTCTC
ACACAACCAAGATACCTTCTGCAAAGGTTGTGCAAGGTATTGGACCGAAGGTGGATCTCTT
AGGAATGTTCTGTGGGAGGAAGCTCAAGAAAGAACAAGAGATCATCTTCATCTTCTTCA
TCAAACATCCTTCAGACAATACCATCTTCACTTCAGATCTAAACCCGCCAATACTCTTC
TCAAACCAATCCATAATAAATCGAAAGGTCATCACAAGATCTCAACTTGTGTGCTTTC
CCAGTCATGCAAGATCAACATCATCATGTCCATATGTCTCAGTTTCTTCAGATGCCT
AAGATGGAGGGAAATGGTAACATAACTCATCAGCAGCAGCCTTCATCATCTTCTCTGTC
TATGGTTCTCTGCTCTCTCTGTTTCAGCTCTTGAACTTTTAAGAACCGAGTTAATGTT
TCTTCAAGATCAGGGATTAACTCATCGTTCATGCCTTCCGGTTCAATGATGGATTCAAAC
ACTGTGCTTTTACTTCTTCAAGGTTTCCAACAATGGTGGATTACAGCCAAGTAATCTC
TCCTTCTCTACCGATCATCAAGGGCTTGGACACAATAGCAACAATAGGTCTGAAGCTCTT
CATAGTGATCATCACCACAAGGTAGAGTTTTGTTTTCCATTTGGGGATCAAATGAAGGAG
CTTTTCATCAAGCATAACACAAGAAGTTGATCATGATGATAATCAACAACAGAAGATCAT
GGAAATAATAATAATAACTCAAGCCCTAATAATGGATATTGGAGTGGGATGTTTC
AGTACTACAGGAGGAGGATCTTCATGGTGA
>G1883 Amino Acid Sequence (domain in aa coordinates: 82-124)
MDATKWTQGFQEMMNVPMEQIMIPNNNTHQPNNTSNARPNTILTSNGVSTAGATVSGVS

NNNNNTAVVAERKARPQEKLNCPRCNSTNTKFCYNNYSLTQPRYFCKGCRRYWTEGGSL
RNVFVGSSSRKNKRSSSSSSSSNIQTIPSSLPDLNPPILFSNQIHNSKSGSSQDLNLLSF
PVMQDQHHHHVHMSQFLQMPKMEGNINITHQQQPSSSSSVYSSSSSPVSALELLRTGVNV
SSRSGINSSFMPSGSMMDSNTVLYTSSGFPTMVDYKPSNLSFSTDHQGLGHNSNNRSEAL
HSDHHQQGRVLFPPFGDQMKELSSSITQEVDDNDNQQKSHGNMNNNNSSPNNGYWSGMF
STTGGGSSW*

>G1884 (1..741)

ATGATGACGTCAATCCATCAGAGCAACACCACCGGCTTTAAACCGCGCGGATCAAGACG
ACGGCGAAGCCACCACGTCAGATCAATAACAAAGAACCATCTCCGGCGACGCAGCCGGTG
CTCAAGTGTCCGAGATGTGATTCACTCAACACCAAATCTGCTACTACAACAACTACAGC
TTGTCTCAGCCACGTCCTACTGCAAGAACTGTCGTCGTTACTGGACACGTGGCGCGGCC
CTCCGTAACGTTCCCATCGGTGGCTCCACTCGAAACAAGAACAAGCCTTGACGCTCCAA
GTCATCTCTTCTCCTTCTGTTTCTCGAACGGGACGTCATCGGCGTCTCGTGAGCTTGTA
AGAAACCATCCATCGACGGCAATGATGATGATGAGTTCTGGTGGATTCTCCGGCTATATG
TTTCCGTTGGATCCTAACTTCAACCTTGCCTCGTCTTCTATCGAGTCTTTGAGTCTTTT
AACCAAGATTTGCACCAGAAGCTTCAGCAACAAGACTCGTCACTTCCATGTTTCTCCAA
GATTCTCTTCCGGTTAACGAGAAAACGGTTATGTTTCAGAACGTAGAGTTGATTCTCTCT
TCGACGGTGACGACGGATTGGGTTTTTCGATAGGTTGCCACTGGAGGAGGTGCAACAAGT
GGCAATCATGAAGATAATGATGATGGGGAGGGTAATTTGGGAAATTGGTTCCATAATGCT
AATAATAATGCTCTGCTCTAA

>G1884 Amino Acid Sequence (domain in AA coordinates: 43-71)

MMTSSHQSNITGFKPRRIKTAKPPRQINNKEPSPATQPVLCPRCDVNTKFCYNNYS
LSQPRHYCKNCRRYWTRGGALRNVPIGGSTRNKNKPCSLQVISSPPLFSNGTSSASREL
RNHPSTAMMMSSGGFSGYMFPLDPNLFNLASSIESLSFNDLHQKLQQQLVTSMLFQ
DSLFPVNEKTVMFQNELIPPSTVTTDWVDFRATGGGATSGNHEDNDDEGNLGNWFHNA
NNNALL*

>G1891 (1..750)

ATGGATAACTTGAATGTTTTCGCAAATGAAGACAATCAAGTGAATGATGTGAAGCCCCCA
CCACCACCACCTCGAGTGTGTGCAAGGTGTGATTCTGATAATACTAAATTTTGTTATTAC
AACAACTACTGTGAGTTTCAGCCACGATACTTCTGCAAGAACTGTCTGATAGATACTGGACT
CATGGTGGGGCTTTAAGAAACATACCAATTGGTGGAAAGTAGTCGTGCCAAACGGGCAAGG
GTAAATCAACCTTCGGTTGCTCGGATGGTTTCTGTTGAGACCCAACGAGGTAAACATCAA
CCTTTCTCTAATGTTCAAGAAAACGTTTCATCTTGTGGATCTTTTGGTGCTTCATCTTCA
TCTTCTGTTGGTGCTGTTGGGAACCTTTTTGGTTCTTTGTATGATATTATCATGGTGGTATG
GTAACAAATTTGCATCCAACCTCGAAGTGTTCGACCAAATCATCGCTTAGCTTTCCATGAT
GGATCATTGAGCAAGACTATTACGATGTTGGGTCCGATAATCTTTTGGTCAACCAACAA
GTTGGTGGCTACGGTTATCAGATGAATCCAGTGGATCAATTCAAGTGGAAACCAGAGCTTC
AACAACACTATGAACATGAATTATAATAACGATAGCACTAGTGAAGTAGCAGAGGATCT
GACATGAATGTGAACCATGATAACAAGAAGATCAGATACCGCAACTCTGTGATTATGCAT
CCTTGTCTCTGGAGAAGGATGGTCCTTGA

>G1891 Amino Acid Sequence (domain in aa coordinates: 27-69)

MDNLNVFANEDNQVNDVKPPPPPRVCARCDSDNTKFCYNNYCEFPQRYFCKNCRRYWT
HGGALRNIPIGGSSRAKRARNVQPSVARMVSVETQRGNQPFQSNVQENVHLVGSFGASS
SSVGAVGNLFGSLYDIHGGMVNHLHPTRTVRPNHRLAFHDGSFEQDYDVGSDNLLVNQQ
VGGYGYHMPVDQFKWNQSFNNTMNMNYNNDSTSGSSRSDMNVDNKKIRYRNSVIMH
PCHLEKDG*

>G1896 (1..951)

ATGTCCTCCCATAGCAATCTCCCTCTCCCAAACAGTTCCTAAACCAGATCACCGTATC
TCCGGTACATCCCAAACCAAGAAACCACCGTCTTCTCCGTAGCTCAAGACCAACAAAC
CTAAAATGCCCTCGTTGCAACTCTCCAAACACAAAGTTCTGTTACTACAACAACTACAGT
CTCTCTCAACCTCGTCACTTCTGCAAACTTGTGCGCGTTACTGGACACGTGGCGGTGCT
CTAAGAAACGTCCCCATCGGTGGTGGTTGCCGGAACCAAAAAATCTATCAAACCTAAT
TCCTCCATGAACACACTTCTCTCTTCTTCTCTCAGAGGTTCTTCTCATCAATCATG
GAAGATTCAACCAATTTCTCCCTCCGACAACAATGGATTTTTCAGCTGGCCGGATTA
TCTCTCAACAAAATGAACGATCTTCAACTTTTGAATAACCAAGAAGTTCTTGATCTTAGG
CCCATGATGTCCTCGGGCCGAGAAAACACACCCGTTGATGTCGGGTGCGGTTTATCCCTA
ATGGGTTTTGGAGATTTCAACAACAACCATTCACCGACGGGGTTCAACCCGCGGAGCA

AGCGACGGAACTTAGCTTCTTCTATAGAGACTTTGAGTTGTTTAAACCAAGATTTACAC
TGGAGGCTTCAGCAACAGAGGATGGCGATGCTTTTTGGTAATTCTAAGGAAGAACTGTT
GTCGTCGAGAGGCCACAACCTATTCTTTATCGGAATCTTGAGATCGTAAACTCATCATCG
CCGTCGTCGCCGACGAAGAAAGGAGATAATCAGACAGAGTGGTATTTTGGTAATAACAGT
GATAATGAAGGAGTGATTAGTAATAATGCTAATACAGGAGGAGGAGGAAGTAATGGAAC
AATGGAATTCAAGCTTGGACTGATCTTAATCATTATAATGCATTGCCTTGA

>G1896 Amino Acid Sequence (domain in aa coordinates: 43-85)
MSSHTNLPSPKPVKPDHRISGTSQTKPPSSSVAQDQQLKCPRCNSPNTKFCYNNYS
LSQPRHFCKSCRRYWTRGGALRNVPFIGGCRKTKKSIKPNSSMNTLPSSSSSQFFSSIM
EDSSKFFPPPTTMDFQLAGLSLNKMNDLQLLNNQEVLDLRPMMSSGRENTPVGVGSLSL
MGFGDFNNHNSPTGFTTAGASDGNLASSIETLSCLNQDLHWRLQQORMAMLFGNSKEETV
VVERPQPILYRNLEIVNSSSPSPTKKGDNQTEWYFGNNSDNEGVISSNNANTGGGGSEWN
NGIQAWTDLNHYNALP*

>G1898 (1..630)

ATGCCGTCGGAACCAACCAACCCGACCCACCAGAGTTCAGCCCTCAACGGCGGCTTAC
CCACCGCCAAATCTGGCTGAGCCTCTTCCTTGTCTCGCTGCAACTCCACCACCACCAAG
TTCTGTACTACAACAACTATAACCTCGCTCAGCCTCGCTACTACTGCAAATCTTGCCGC
CGTTACTGGACTCAAGGTGGTACACTCCGTGACGTCCCCGTGGTGGTGAACCTCGTCGA
AGCTCCTCAAAACGTCACCGTTCTTTCTCCACCACTGCCACCTCTCTTCTCTCTTCT
TCCGTATCACCACCAGACACAAGAACCAGCCACGACTGAAGCGAGTCAAACCTAAGGTT
ACTAATTTAATTTCAGGTCATGGAAGCTTTGCTTCTCTGTTAGGTTTAGGAAGTGGAAT
GGTGGGTTGGATTACCGGTTTGGGTACGGGTACGGGCTTGAGGAGATGAGTATTGGGTAT
CTTGAGATTCTTCCGTAGGAGAGATTCCGGTGGTTGATGTTGTGGTGGTGACACGTGG
CAGATTGGGGAGATTGAAGGTAAAGTGGAGGAGACAGTTTGATATGGCCTGGTCTTGAG
ATCTCAATGCAAACCAACGATGTTAAGTGA

>G1898 Amino Acid Sequence (domain in AA coordinates: 31-59)
MPSEPNQTRPVRVQSTAAAYPPPNLAEP LCPRCNSTTKFCYNNYNLAQPRYCKSCR
RYWTQGGTILRDVPVGGGTRRSSSKRHSFSTTATSSSSSSSVITTTQEPATTEASQTKV
TNLISGHGSFASLLGLGSGNGGLDYGFGYGYGLEEMSIGYLGDSVGEIPVVDGCGGDTW
QIGRIEKGSGGSLIWPGLISMQTNDVK*

>G1902 (1..615)

ATGCAGGATCCAGCAGCATATTACCAGACGATGATGGCGAAGCAACAACAACAACAA
CCACAGTTTGCAGAGCAAGAACAGTTAAAGTGTCTCGTTGTGACTCACCAAACACTAA
TTCTGTACTACAACAACCTACAATCTCTCACAGCCTCGTCACTTTTGCAAAGCTGTCGT
CGTTACTGGACTAAAGGCGGCGCTCTCCGTAACGTTCCCGTCGGTGGTGGTTCTCGTAAG
AACGCAACCAAACGATCCACTTCTTCTTCTTCTTCTCTCTCTCTCTCTCTCTCTCTCT
CAAAACAAGAAGACGAAAAACCCGGATCCGGATCCTGATCCACGTAATTCTCAAAAACCG
GATTTGGATCCGACCCGGATGCTTTACGGGTTTCCGATCGGTGACCAAGACGTGAAGGGT
ATGGAGATTGGTGAAGCTTTAGCTCGTTGTTGGCGAATAATATGCAGCTTGGTCTTGGA
GGAGGAGGATCATGCTTGACGGGTCGGGTTGGGATCATCCGGGTATGGGTTTGGGTTTG
AGGAGAACCGAACCGGTAATAATAATAATAACCCATGGACCGATCTGGCTATGAACAGA
GCGGAGAAAACTGA

>G1902 Amino Acid Sequence (domain in AA coordinates: 31-59)
MQDPAAYYQTMMAKQQQQPQFAEQEQLKCPRCDSPTKFCYNNYNLSQPRHFCKSCR
RYWTKGGALRNVPVGGGSRKNATKRSTSSSSSASSPSNSSQNKKTKNPD PDPDRNSQKP
DLDPTRMLYGFPIGDQDVKGMEIGGSFSSLLANNMQLGLGGGIMLDGSGWDHPPMGLGL
RRTEPGNNNNPWTDLAMNRAEKN*

>G1904 (1..924)

ATGCAAGATATTCATGATTTCTCCATGAACGGAGTTGGTGGTGGGGGAGGAGGAGGG
AGGTTTTTTCGGTGGAGGAATCGGCGGCGGAGGAGGTGGTGATCGAAGGATGAGAGCTCAT
CAGAACAATATACTTAACCATCATCAATCTCTCAAGTGTCTCGTTGTAATCTCTTAAC
ACAAAGTTCTGTACTACAACAATTACAATCTTTCTCAGCCTCGTCACTTTTGCAAGAAC
TGTCGTCGTTACTGACTAAAGGTGGTGTCTCCGTAACGTTCCCGTCGGAGGTGGTTGC
CGGAAAGCTAAACGTTCCGAAAACAAAACAGGTTCCGTCGTCGTCATCAGCCGACAAACCA
ACGACGACGCAAGATGATCATCAGTGGAGGAGAAATCGAGTACAGGATCTCACTCTAGC
AGCGAGAGCTTCTCTCACCGCTTCTAACTCTACCACCGTCGCCGCCGTCTCCGTCACC
GCGGCGGCGGAAGTTGCTTCGTGCTTATTCAGGTTTTGATATGCCTAATATGAAAATT

TACGGTAACGGGATCGAGTGGTCGACGTTACTTGGACAAGGCTCATCGGCCGGTGGTGT
TTCTCGGAGATCGGTGGTTTTCCGGCGGTTTCAGCTATTGAACTACACCGTTTGGATTC
GGGGGTAAATTCGTAATCAAGATGATCATCTGAAGTTAGAAGGTGAACTGTACAGCAG
CAACAGTTTGGAGATCGAACGGCTCAGGTTGAGTTTCAAGGAAGATCTTCGGATCCGAAT
ATGGGATTTGAACCGTTGGATTTGGGGAAGTGGCGGTGGAGATCAAACACTGTTTGATT
ACCAGTACCGTTGATCATGCATACTGGAGTCAAAGTCAATGGACGTCGTCTGACCAAGAT
CAGAGTGGTCTCTACCTTCCTTGA

>G1904 Amino Acid Sequence (domain in aa coordinates: 53-95)
MQDIHDFSMNGVGGGGGGGRFFGGGIGGGGGDRRMRAHQNNILNHHQSLKCPRCNSLN
TKFCYYNNYLSQPRHFKNCRRYWTGGVLRNVPVGGGCRKAKRSKTKQVPSSSSADKP
TTTQDDHHVEEKSSTGSHSSSESSSLTASNSTTVAASVTAAAEVASSVIPGFDMPNMKI
YNGIEWSTLLGQSSAGGVFSEIGGFPAVSAIETTPFGFGGKFNQDDHLKLEGETVQQ
QQFGDRTAQVEFQGRSSDPNMGFEPLDWGSGGGDQTLFDLTSTVDHAYWSQSQWTSSDQD
QSGLYLP*

>G1906 (1..795)
ATGGTGAACGTGCTCGGATCGCAAAAGTCCCATTCCTGAAGCAGCTCTAAATTGCCCT
AGATGTGACTCAACCAATACTAAGTTCTGTACTTCAATAACTATAGCCTTACTCAACCT
CGCCATTTCTGCAAAACATGTCGTCGCTATTGGACACGTGGCGGTTCTTGGAGGAATGTT
CCTGTTGGAGGAGGCTTTAGGAGGAACAAGAGAAGCAAATCCAGATCGAAATCTACGGTC
GTGGTCTCGACTGATAATACTACTAGTACTTCATCACTTACTTCTCGCCCAAGTTACTCA
AACCCTAGCAAGTTTCATAGCTACGGTCAAATCCCGAGTTTAATTCCAACCTGCCCATC
TTGCCTCCTCTCCAAAGCCTTGAGATTACAATCAAGCAACACTGGATTAGATTTTGGT
GGAACCTCAAATAAGCAACATGATAAGTGGTATGAGTTCTAGTGGTGGGATCTTGGATGCA
TGGAGAATACCTCCATCACAACAAGCTCAGCAATCCCTTTCTTGATCAACACTACCGGA
TTGGTGCATCTTCAAACGCGTTATATCCATTACTAGAAGCGGGGTTAGCGCCACGCAA
ACAAGAAATGTGAAGGCGGAAGAGAATGATCAGGATCGGGGTAGGGATGGGGATGGAGTG
AATAACTTATCAAGAACTTTTTGGGTAATATCAACATAAACTCAGGCAGGAACGAGGAA
TACACATCATGGGGAGGTAACAGTCTTGGACCGGTTTCACCTCCAACAACCTCAACAGGC
CATCTCTCATTCTAA

>G1906 Amino Acid Sequence (domain in AA coordinates:19-47)
MVERARIAKVLPEALNCPKCDSTNTKFCYFNYSLTQPRHFKTCRRYWTRGGSRLNV
PVGGGFRNRKRSKRSKSTVVVSTDNTTSTSSLTSPSYSNPSKFHSYQIPEFNSNLPI
LPPLQSLGDYNSNTGLDFGGTQISNMISGMSSSGIILDAWRIPPSQQAQQFPFLINTTG
LVQSSNALYPLLEGGVSATQTRNVKAEENDQDRGRDGDGVNLSRNFLGNININSGRNEE
YTSWGGNSSWTGFTSNNSTGHLSF*

>G1913 (1..744)
ATGGAGAGAGCAGAGCCCTTGACATCATCGTTTTATATGGCGGCCAAACGCAACGCAAAAC
GCGGAGATCACGCCGAGTTGTCCAAGATGTGGATCCTCTAACACAAAGTTCTGTACTAC
AACAACTATAGCCTCACTCAGCCTCGCTACTTCTGCAAAGGCTGCCGAGATATTGGACC
AAAGGTGGTTCCCTCCGCAATGTTCTGTAGGCGGTGGCTGTGCGAAATCCCGCGCCCC
AAATCATCTTCTGGTAACAATACTAAAACTAGCCTAACCCTAATTCTGGCAACCCCGGT
GGTGGTTACCAAGCATCGATCTTGCTCTTGTTTACGCCAATTTCTTGAATCCAAAGCCT
GACGAATCTATACTACAAGAAAATTGCGACTTAGCCACTACGGATTTTTTGGTAGATAAT
CCTACCGGCACTTCCATGGACCTTCATGGAGTATGGACATCAATGATGGTCATCATGAT
CATTATATTAATCCGGTGGAACACATTGTGGAGGAATGTGGTTATAATGGCTTGCCCTCCA
TTTCTGGTGAAGAGCTTCTCTCTTTAGACACTAATGGTGTGGTCTGATGCTTTGTTG
ATTGGTCATAACCATGTAGACGTTGGCGTGAATCCGGTTCAGGCTGTACACGAACCGGTG
GTTCAATTCGCTGAAGAATCCAATGATCCACCAATCTCTTGTTTGGAAGTTGGAGCCCT
TTTGATTTCACTGCCGATGGATGA

>G1913 Amino Acid Sequence (domain in AA coordinates: 27-55)
MERAELTSSFIWRPNANANAEITPSCPRGSSNTKFCYNNYSLTQPRYFCKGCRRYWT
KGGSLRNVPVGGGCRKSRPKSSSGNNTKSLTANSGNPGGSPSIDLALVYANFLNPKP
DESILQENCDLATDFLVDNPTGTSMDPSWSMDINDGHHDHYINPVEHIVEECGYNGLPP
FPGEELLSLDTNGVWSDALLIGHNHVDVGVPVQAVHEPVVHFADESNDSTNLLFGSWSP
FDFTADG*

>G1914 (1..945)
ATGGAGAGATACAAGTGTAGATTTTGCTTCAAGAGCTTCATCAATGGAAGAGCTTTAGGT

GGTCACATGAGATCTCACATGCTTACTCTTTCTGCAGAACGTTGTGTAATAACTGGTGAA
GCAGAAGAAGAGTAGAGGAACGGCCGAGTCAACTCTGTGACGACGACGACGATACCGAG
TCCGATGCTTCTTCTTCTTCTGGTGAGTTTGATAATCAAAAGATGAATCGTCTTGATGAT
GAATTGGAGTTTGATTTTCGCTGAAGACGACGACGTTGAAAGTGAAACCGAGTCGTCCAGG
ATTAACCCAACCTCGGCGACGATCTAAACGAACTCGGAAACTTGGATCGTTTGATTTTCGAC
TTTGAGAAGCTAACAACGAGCCAACCCAGTGAGTTAGTGGCCGAGCCAGAGCATCACAGC
TCAGCTTCTGATACAACAACGGAGGAAGATCTCGCCTTTTGTCTCATTATGCTGTCCAGA
GACAAATGGAAGCAACAGAAGAAGAAGCAACGTGTAGAAGAAGATGAGACAGATCAT
GACAGTGAAGATTACAAATCAAGCAAGAGCAGAGGGAGATTCAAGTGTGAGACTTGTGGT
AAAGTGTTTAAATCGTATCAAGCATTAGGAGGACACAGAGCAAGCCACAAGAAGCAAG
GCATGCATGACGAAAACAGAGCAAGTTGAAACAGAGTACGTTCTTGGAGTAAAGGAGAAG
AAAGTTCATGAATGTCCGATCTGTTTTAGGGTTTTTACTTCAGGGCAAGCACTTGGAGGT
CATAAGAGATCTCACGGAAGTAACATCGGAGCAGGAAGAGGATTGTCAGTAAGTCAAATT
GTCCAAATCGAAGAAGAAGTATCAGTGAACAGAGGATGATTGATCTTAATCTTCCTGCA
CCTAATGAAGAAGATGAAACTTCTTTGGTGTGATGAATGGTGA

>G1914 Amino Acid Sequence (domain in AA coordinates:195-216, 245-266)

MERYKCRFCFKSFINGRALGGHMRSHMLTSLAERCVITGEAEVEVEERPSQLCDDDDTE
SDASSSSGEFDNQKMNRLDDELEFDFAEDDDVESETESSRINPTRRRSRKTRKLGSPDFD
FEKLTTSPSELVABPEHHSSASDTTTEEDLAFCLIMLSRDKWKQKKKKQORVEEDET
DSEYDKSSKSRGRFKCETCGKVFKSYQALGGHRASHKKNKACMTKTEQVETEVVLGVKEK
KVHECPICFRVFTSGQALGGHKRSHGNSNIGAGRGLSVSQIVQIEEEVSVKQRMIDLNLPA
PNEDETSLVFDEW*

>G1925 (1..945)

ATGGAAGAAAATCTTCTCCGGGGTTTCTAGATTTCATCCTACAGACGAGGAGCTCATAACG
CATTATCTATGTGCGAAAGTCTCCGATATAGGATTACCGGTAAAGCTGTCTGTCGACGTT
GATCTCAACAAGTGTGAACCTTGGGATTTGCCAGCCAAGGCTTCAATGGGAGAGAAAGAG
TGGTATTTCTTTCAGCCAAAGGGATCGGAAATATCCAACCGGTTTAAAGAACAAACCGGGCA
ACAGAAGCTGGTTACTGGAAAACCAACCGGGAAAGATAAAGAAATATACCGAAGTGGAGTG
TTGGTTGGGATGAAGAAAACCTAGTTTCTACAAAGGAAGAGCTCCCAAAGGTGAGAAA
AGCAATTGGGTTATGCATGAGTACAGGCTTGAGAGCAAAACAACCTTCAACCCACGAAT
AAGGAGGAATGGGTAGTGTGTAGGGTTTTTCGAAAAGAGCACGGCAGCAAAGAAAGCACAA
GAACAACAACCTCAATCTTCTCAACCATCTTTTGGATCTCCATGCGATGCAAACCTCATCA
ATGGCAAATGAGTTTGAAGATATTGATGAGCTTCCGAATCTGAATCAAACCTCATCAACC
ATCGATTACAATAATCATATCCATCAATATTCGCAACGCAATGTTTACTCAGAAGACAAC
ACAACAAGTACGGCTGGTCTCAACATGAACATGAACATGGCTAGTACTAATCTTCAGTCT
TGGACAACAAGTCTCCTTGGTCCGCCTTTATCTCCAATCAACTCTTTGTGCTCAAGGCT
TTCCAAATCAGGAACCTTTATAGTTTCCCAAAGAGATGATCCCAGTTTCAATCATTCT
TCTCTTCAACAAGGAGTCTCCAATATGATCCAAATGCTTCAAGTTCGTCTCAAGTGCAA
CCCCAACCGCAAGAGGAAGCGTTAATATGGACTCCATATGGTGA

>G1925 Amino Acid Sequence (conserved domain in AA coordinates:6-150)

MEENLPPGFRFHPTDEELITHYLCRKVSDIGFTGKAVVDVLDNKCEPWLPAKASMGKEKE
WYFFSQDRKYPTGLRTNRATEAGYWKTTGKDKEYRSGVLVGMKKTLLVFKGRAPKGEK
SNWVMHEYRLSKQPFNPINKKEWVVCRVFEKSTAAKKAQEQPQSSQPSFGSPCDANSS
MANEFEDIDELPNLNSNSSTIDYNHIIHQYSQRNVYSEDNTTSTAGLNMNMMASTNLQS
WTTSLGLPPLSPINSLLLKAFQIRNSYSFPKEMIPSFNHSSLQQGVSNMIQNASSSSQVQ
POPQEEAFNMDSIW*

>G1929 (1..366)

ATGTGTAGAGGCTTGAATAATGAAGAGAGCAGAAGAAGTGACGGAGGAGGTTGCCGGAGT
CTCTGCACGAGACCGAGTGTTCCGGTAAGGTGTGAGCTTTGCGACGAGACGCCTCCGTG
TTCTGTGAAGCGGACTCGGCGTTCTCTGTAGAAAATGTGACCGGTGGGTTTCATGGAGCG
AATTTTCTAGCTTGGAGACACGTAAGCGCGGTGCTATGCACTTCTTGTGAGAACTCACG
CGCCGGTGCCTCGTCGGAGATCATGACTTCCACGTTGTTTTACCGTCGGTGACGACGGTC
GGAGAAACCACCGTGGAGAATAGAAGTGAACAAGATAATCATGAGGTTCCGTTTGT
CTCTGA

>G1929 Amino Acid Sequence (domain in AA coordinates:31-53)

MCRGLNNEESRRSDGGGCRSLCTRPSVPVRCELCDGDASVFCEADSAFLCRKCDRWVHGA
NFLAWRHVRRVLCTSCQKLTRRCLVGDHDFHVVLPSVTTVGETTVENRSEQDNHEVPFVF

L*

>G1930 (76..1077)

ATTCACATTACTAATCTCTCAAGATTTCAAAATTTTCTGTGATTTTCTCTCAGTTTCTT
ATTTTCGTTTCATAACATGGATGCCATGAGTAGCGTAGACGAGAGCTCTACAACCTACAGAT
TCCATTCCGGCGAGAAAGTCATCGTCTCCGGCGAGTTTACTATATAGAATGGGAAGCGGA
ACAAGCGTGGTACTTATTAGAGAACCGGTGTGGAAGTCGAAGTCGAAGCCGAATCAAGA
AAGCTTCTTCTTCAAGATTCAAAGGTGTTGTTCTCAACCAAATGGAAGATGGGGAGCT
CAGATTTACGAGAAACATCAACGCGTGTGGCTTGGTACTTTCAACGAGGAAGACGAAGCA
GCTCGTGCTTACGACGTGCGGGCTCACCGTTTCCGTGGCCGCGATGCCGTTACTAATTTT
AAAGACACGACGTTTCAAGAAGAGGTTGAGTTCTTAAACGCGCATTGGAATCAGAGATC
GTAGATATGTTGAGAAAAACACTTACAAAAGAGGTTAGACCAAAGGAAACGTAACCGT
GACGGTAACGGAAAAGAGACGACGGCGTTTGTCTTGGCTTCGATGGTGGTTATGACGGGG
TTTAAACGGCGGAGTTACTGTTTGAAGAACGGTAACGCCAAGTGACGTGCGGAAACTA
AACCCTTTAGTTATACCAAAACACCAAGCGGAGAAACATTTCCGTTACCGTTAGGTAAT
AATAACGTCTCCGTTAAAGGTATGCTGTTGAATTTCAAGACGTTAACGGGAAAGTGTGG
AGGTTCCGTTACTCTTATTGGAATAGTAGTCAAAGTTATGTGTTGACCAAAGGTTGGAGT
AGATTCGTTAAAGAGAAGAGACTTTGTGCTGGTGATTTGATCAGTTTTAAAGATCCAAC
GATCAAGATCAAAAATCTTTATCGGGTGGAAATCGAAATCCGGGTTGGATCTAGAGACG
GGTCGGGTTATGAGATTGTTTGGGGTTGATATTTCTTTAAACGCCGTCGTTGTAGTGAAG
GAAACAACGGAGGTGTTAATGTGCTCGTTAAGGTGTAAGAAGCAACGAGTTTTGTAATAA
CAATTTAACACTTGGGAAAGAAAAAAAGCTTTTGTATTTAATTTCTCTTCAACGTTA
ATCTTGCTGAGATTA

>G1930 Amino Acid Sequence (domain in AA coordinates: 59-124)

MDAMSSVDESSTTDSIPARKSSSPASLLYRMGSGTSVVLDSENGVEVEVEAESRKLPS
RFGVVPQPNRGRWGAQIYEKHQRVWLGTFFNEEDEAARAYDVAAHRFRGRDAVTNFKDTTF
EEVEFLNAHSKSEIVDMLRKHTYKEELDQKRNRDNGNKETAFALASMVMTGFKTAE
LLFEKTVTPSDVGLNRLVIPKHQAEEKHFPPLPLGMNNVSVKGMLLNFEDVNGKVWRFRYS
YWNSSQSYVLTKGWSRFVKEKRLCAGDLISFKRSNDQDQKFFIGWKSGLDLETGRVMR
LFGVDISLNAVVVVKETTEVLMSSLRCKKQRL*

>G195 (51..1031)

TTTTCTTTTCTTTCTTTTGGTTTAAAGTTTTTCTCTTTGTTCTTCGTCATGTCTCATG
AAATCAAAGATCTTAACAACCTACTACACTTCATCGTATAATCATTACAATATCAACA
ACCAAAATATGATTAATCTCCCTTACGTTTCTGGTCCATCTGCTTATAATGCAAACATGA
TCTCATCATCACAAGTAGGTTTTGATCTACCCTCGAAGAACTTGAGTCTCTCAAGGAGCCT
TCGAGTTGGGTTTCGAGCTTTCTCCATCTTCTTCTGACTTTTTTAATCCTTCCCTCGATC
AAGAGAACCGTTTGTATAATGCTTATAATTATAATAGTAGTCAAAAGAGTCATGAAGTTG
TCGGTGATGGTTGTGAACCATTAAGAGTGAAGTTAGGGTTTCAGCATCTCCTTCTTCAA
GTGAGGCCGATCATCATCAGGAGAAGATTCCGGCAAGATCCGGAAGAAAAGAGAAGTTC
GCGATGGAGGAGAATGATCAACGCTCTCAGAAAGTAGTTAAAACAAAGAAGAAAGAGG
AGAAGAAAAAGAGCCACGAGTCTCGTTCATGACTAAGACCGAAGTTGATCATCTCGAAG
ACGGCTATCGTTGGAGAAAGTATGGCCAAAAGCAGTCAAAAACAGTCTTATCCGAGGA
GTTACTATAGATGCACGACTCAGAAGTGCAACGTGAAGAAGAGAGTGGAGAGATCTTACC
AAGACCCAACGGTTCGTCATCACAACCTACGAGAGTCAACACAACCATCCGATCCCGACCA
ATCGTCCGACAGCAATGTTCTCTGGAACACCGCATCTGATTATAACCCATCATCGTCTC
CAATATTTCCGATCTCATCATCAATACTCCAAGAAGCTTCTCAAATGATGATCTCTTCC
GTGTGCCATACGCTAGTGTGAACGTGAACCCCTAGTTATCATCAACAGCAACATGGATTTT
ATCAACAGGAGAGTGAGTTTCGAGCTCTTGAAGGAGATGTTTCTCGGTTTTCTTCAAAC
AAGAGCCTTGATGATATAATATAATATAGAAACAATTTTTTTCTGCTAAGAAATATAGA
ACAAAACCTGGATGCATAATAAGTGATGATAGTGTTATTTATTTTTTGCATGTATATATT
ATACATGTTTTGTTAACTAGCTATAGGATATACTGGTAGTAATTAAGCATAAATATGGAG
CCCTTCGACTTATTACAATAATTTTTGGTATGGAAGAAANTTNGNTACATGCCTGCCTTTT
NNNTTNNNG

>G195 Amino Acid Sequence (domain in AA coordinates: 183-239)

MSHEIKDLNNYHYTSSYNHYNINNQNMINLPYVSGPSAYNANMISSQVGFDLPSKNLSP
QGAFELGFELSPSSSDFFNPQLDQENGLYNAYNINSSQKSHEVVGDGCATIKSEVRVSAS
PSSSEADHHPGEDSGKIRKKREVRDGGEDDQRSQKVVKTKKKEKKKEPRVSFMTKTEVD
HLEDGYRWRKYGQKAVKNSPYPRSYRCTTQKCNVKKRVERSYQDPTVVITTYESQHNP

IPTNRRRTAMFSGTTASDYNPSSSPIFSDLIINTPRSFSNDDLFRVPYASVNVNPSYHQOQ
HGFHQESEFELLKEMFSPVFFKQEP*
>G1954 (196..1440)
ATTTATGACTTCTCAATACAAAAGCTCCCCTCACTTTTTTAAGTTTTGTCTTCTCTAAT
CCGTCTTCTTCTACTATCTTGCATGTCTTGCGTCTTTTATATACATCTCTCGTAAACCCCT
AGCAAATCATACAAAGGTCAAGAAGCTTGACCTTCATTAGACTTAAGCAGTTTATAATCAA
CTACCACGAATAGCAATGGATAAAGATTACTCGGCACCAAACCTTCTTAGGTGAATCCTCA
GGCGGTAACGATGATAACAGCTCTGGTATGATAGACTATATGTTCAATAGAAACCTTCAA
CAACAACAAAAGCAATCGATGCCACAACAGCAGCAACATCAACTCTCTCCTTCCGGATTT
GGAGCAACACCCCTTTGATAAAATGAACCTTCTCTGATGTGATGCAGTTTGCGGACTTCGGT
TCGAAACTTGCGTTGAACCAGACCAGAAACCAAGACGATCAAGAAACCGGGATTGACCCC
GTTTATTTCTTGAAGTTCCTGTCTTGAACGACAAAATAGAGGACCATAACCAAACCCAA
CATCTCATGCCTTCTCATCAGACGTCTCAAGAAGGAGGTGAGTGTGGAGGAAACATAGGC
AATGTGTTTTCTTGAAGAAAAAGAAGATCAAGACGATGACAACGACAACAACCTCCGTGCAA
CTACGTTTTTATTTGGAGGAGAAGAAGAAGATAGGGAGAACAAGAATGTTACGAAAAAGGAG
GTGAAGAGCAAGAGGAAGAGAGCTAGAACGAGCAAGACCAGCGAAGAAGTGGAAAGCCAA
CGGATGACTCATATCGCGGTGCAAGAAACCGTAGGAAGCAAATGAATGAGCATCTTCGT
GTCCTTAGATCTCTCATGCTTGGCTCCTACGTTCAAAGGGGAGACCAAGCGTCAATCATA
GGAGGAGCAATAGAGTTTGTGAGAGAGCTCGAGCAACTCCTACAATGTCTTGAATCACAG
AAGCGTCTGAAGAATCTTAGGAGAAACCGGTAGGGACATGACAACGACAACGACTTCTTCT
TCTTCTCCATAACTACGGTAGCGAACCAGCACAACCGCTCATATTACGGGAAATGTA
ACCGAGCTAGAGGGCGGAGGAGGCTTTCGGGAGGAGACTGCGGAGAACAAGTCGTGCTTG
GCTGACGTGGAGGTGAAGCTGCTAGGGTTTGACGCCATGATCAAGATACTTTCAAGAAGA
AGGCCGGGACAGCTGATTAAGACTATAGCTGCTTTGGAGGATCTTCATCTCTCTATTCTT
CACACTAACATCACTACCATGGAACAACCGTCTCTACTCCTTTAATGTCAAGATAACA
AGTGAAACGAGGTTTACGGCAGAAGACATAGCAAGTTCATCCACAGATATTTAGTTTC
ATTCTATGCAAATACCAATATCTGGAAGCTCTAACCTGGGAAATATTGTGTTTACTTGA
AAATCATCACACGGCGACAACCTTTGTACACTGGTGAAGATTACGTACGTAATAATCTCT
ACATATTGGGTTTTATTCTCCAAGCATTGTGAAGAGTGTGTTAAGTTAAAGGGAGTGTGTA
CTTTATTTTTTTTGGGGCTTTTTTTCATGCAATTTAAATTTTAGTGATGATTGTGTCGCTTG
TAATGTTAGAATCTGTTGTTGTGATTTCTGCTGCTTTGATTGTAGGTTTTGAACAAGCG
GTTTAGAATGCTAAACCACTTATTTACTTGAAATAACTTTTTTTCACAAAAAATAAAAA
AAGAAAAAA
>G1954 Amino Acid Sequence (domain in AA coordinates:187-259)
MDKDYSAFNFLGESSGGNDNSSGMIDYMFNRNLQQQKQSMPPQQQHQLSPSGFGATPF
DKMNFSDVMQFADFGSKLALNQTRNQDDQETGIDPVYFLKFPVLNDKIEDHNQTQHLMP
HQTSQEGGECGNIGNVFLKEKEDQDDNDNNSVQLRFIGGEEEDRENKNVTKEVKS
KRARTSKTSEEVEFSQRMTHIAVERNRRKQMNHLRLSLMPGSYVQRGDQASIIIGGAIE
FVRELEQLLQCLSQRRRIILGETGRDMTTTTSSSSPITTVANQAQPLIITGNVTELEG
GGGLREETAENKSLADVEVKLLGFDAMIKILSRRRPGQLIKTIAALEDLHLSILHTNIT
TMEQTVLYSFNVKITSETRTAEDIASSIQQIFSFHANTNISGSSNLGNIVFT*
>G1958 (107..1336)
GTACCGTCGACCGATTATCCCAAGAGGAGAATCCTCATAATCATTTTCTCCGATTTCGAT
TCGTCTTCTTGGTCTTGATGCTTTCATGAATTTCTAGGACAACAATGGAGGCTCGTCC
AGTTCATAGATCAGGTTTCGAGAGACCTCACACGCACTTCTTCAATCCCATCTACACAAA
ACCTTACCAGTAGAGATAGTTCATGAGATCAGATAACAACAGTCAGTTAATGTCTAG
ACCATTAGGACAAAACCTACCATTTACTTTCATCTAGTAACGGTGGAGCTGTTGGACATAT
ATGTTCTTCTTCATEATCTGGTTTTGCAACCAATCTCCATTACTCAACTATGGTATCTCA
TGAGAAACAACAACACTACACAGGAAGCAGCAGTAATAATGCTGTGCAGACACCAAGCAA
CAACGATAGTGCTTGTTGTGTCATGATTCATTGCCAGGAGGGTTTCTTGACTTCCATGAAAC
CAACCCGGCGATTCAAACAACCTGTGATGAGGATGGTGGCATTGCGGCTGCTTTTGA
TGACATTCAAAAACGAAGTATTGGCATGAATGGGCTGACCATTGATCACTGATGATGA
TCCTTTGATGTCTACTAAGTGAATGATCTCTTGCTTGAACAATAATCCAAATTCAGATT
AAAGGACAGAAAGCACTGCAAAATTCGCAACCTCAGATTGTTTTCAGCAGCAACCTTCTCC
GTCTGTGGAATTTGCGACCTGTTAGCACAACATCTTCAAACAGCAATAACGGAACGGGCAA
GGCAGCAATGCGTTGGACGCCAGAGCTTACGAGGCTTTTGTGAGGCTGTCAACAGTCT
TGGCGGTAGTGAAAGAGCTACTCTAAAGGGGTACTGAAGATTATGAAAGTTGAAGGCTT

GA CTATATATCATGTTAAAAGCCATTTACAGAAATATAGGACAGCTAGATATCGGCCAGA
ACCATCAGAAACTGGTTCGCCAGAAAGGAAGTTGACACCGCTTGAACATATAACATCTCT
TGATTTGAAAGGTGGGATAGGTATTACAGAGGCTCTACGACTTCAGATGGAAGTACAGAA
GCAACTCCATGAGCAGCTCGAGATTCAAAGAAACCTGCAACTCCGAATAGAAGAACAAGG
CAAGTACCTGCAAATGATGTTTCGAGAAGCAAACTCTGGTCTTACCAAAGGGACAGCCTC
AACATCAGATTCCGCAGCCAAATCTGAACAAGAAGACAAGAAGACTGCTGATTCTGAAGGA
GGTTCAGAGAAGAAGAACAGGAAATGTGAGGAAGTGAATCTCCACAGCCAAAGCGTCC
CAAAATCGATAATTGAAAGTATTGGTCTTTTGTGAGATAATCTCGGAGTTTCAGAGTTAA
CAGTGATAGAGAGAACGAGCTCTTATCTTGAGGTTCTTCAGGACTTCTCTCGCGGCCGCT
CTAG

>G1958 Amino Acid Sequence (domain in AA coordinates: 230-278)

MEARPVHRSGSRDLRTSSIPSTQKPSPVEDSFMRSDNNSQLMSRPLGQTYHLLSSSNGG
AVGHICSSSSSGFATNLHYSTMVSHKQHYTGSSSNNVQTPSNND SAWCHDSLPGGFL
DFHETNPAIQNNCQIEDGGIAAFDDIQKRSDWHEWADHLITDDDPMLSTNWNDDLLETN
SNSDSKDQKTLQIPQPIVQQPSPSVELRPVSTSSNSNNGTGKARMRWTPELHEAFVE
AVNSLGGSERATPKGV LKIMKVEGLTIYHVKSHLQKYRTARYRPEPSETGSPERKLTPL
HITSLDLKGGIGITEALRLQMEVQKQLHEQLEIQRLQLRIEEQKYLQMMFEKQNSGLT
KGTASTSDSAKSEQEDKKTADSKEVPEEETRKC EELESPQPKRPKIDN*

>G196 (111..1421)

TCGACATCAGATTTCTCTCACGGATTCTTAATCATTTTTATTATATTGGATATTGCTA
ATTTCTCCCGTGTATAAATCTCATATAAACACGCATCATACATATATATTATGTGCAGCG
TCTTTGAGTTTCAAGACATGGACAACCTCCAAGGAGATCTAACAGACGTCGTACGAGGAA
TAGGATCAGGCCACGTGTCCACATCTCCTGGACCACCGGAAGGTCCATCTCCGAGCAGCA
TGTCTCCGCCGCCAACATCAGATCTCCACGTGGAATTTCCCTCCGCCGCTACTTCTGCCA
GCTGTCTCGCAAATCCCTTCGGAGACCCGTTCTGTAAGCATGAAGGATCCTCTCATCCACC
TCCCGGCCAGCTACATCTCCGGCGCCGGTGATAATAAAAGCAACAAAAGTTTTGCAATCT
TTCCAAAGATTTTTGAGGATGATCATATTAAGAGTCAATGCAGTGTCTTCCCAAGAATTA
AGATCTCGCAAAGTAACAATATCCACGATGCCCTCCACGTGTAATTCTCCGGCCATAACCG
TCTCCTCTGCCCGCTAGCAGCTTCGCCGTGGGGCATGATCAACGTTAATAACCACTAACA
GTCCAAGAAACTGTTTACTTGTGCGATAATAATAACAACACGTCATCATGCTCACAGGTTT
AGATCTCTTCTTCCCTCGGAATCTCGGAATTAAGAGAAGGAAGAGCCAGGCAAAGAAAG
TGGTGTGCATACCGGCTCCAGCCGCTATGAACAGCCGGTCCAGTGGAGAAGTTGTTCCGT
CTGATCTATGGGCTTGGCGAAAGTACGGTCAAAAACCTATCAAAGGTTCTCCTTATCCAA
GGGGTTACTACAGATGTAGCAGCTCAAAAGGTTGTTTCAGCTAGGAAACAAGTCGAACGTA
GCCGCACTGATCCAAACATGTTAGTCATTACTTACACCTCTGAGCATAACCAACCCATGGC
CTACTCAACGCAACGCTCTCGCAGGTTCCACTCGTTCCTCTTCTCCTCCTCTTTAAACC
CTTCTTCCAAATCCCTCAACCGCAGCCGCCACTACTTCTCCCTCATCCAGAGTTTTTCCAA
ACAACAGCAGCAAAGACGAACCCAATAACTCCAACCTTGCCCTTCTTCCACTCATCCTC
CTTTTGACGCCGCCGAATTAAGGAGGAGAACGTGGAAGAGCGTCAGGAAAAGATGGAGT
TCGATTATAATGACGTTGAAAATACCTATAGACCGGAGTTGTTGCAAGAGTTTCAACATC
AGCCGGAGGATTTCTTTGCCGATCTCGACGAGCTTGAGGGAGATTCTTTGACTATGTTGC
TCTCTCACAGTAGCGGCGGAGCAACATGGAAAACAAAACGACGATTCCAGACGTTTTTA
GTGATTTCTTTGACGACGACGAGTCTCAAGGTCGTTATAAATATTGTTGTTAATGTATA
CATAGAAATGAAATTATTCATGTAATTCGTTTTGTGTTAAATGACGGTATTTGCCTTTGC
A

>G196 Amino Acid Sequence (conserved domain in AA coordinates:223-283)

MCSVFQDMDNFQGDLDVVRGIGSGHVSPSPGPPEGPSPPSSMSPPPTSDLHVEFPSAA
TSASCLANPFGDPFVSMKDPLIHL PASYISGAGDNKSNKSF AIFPKIFEDDHIKSQCSVF
PRIKISQSNNIHDASTCNPAITVSSAAVAASPWGMINVNTNSPRNCLLV DNNNNNTSSC
SQVQISSPRNLGIKRRKSQAKKVVCIPAPAAMNSRSSGEVVP SLDLWAWRKYGQKPIKGS
PYPRGYRCSSSSKGCSARKQVERSRTDPNMLVITYTSEHNHPWPTQRNALAGSTRSSSSS
SLNPSSKSSATAATSPSSRVFQNNSSKDEPNNSNLPSSSTHPPFDAAAIKEENVEERQE
KMEFDYNDVENTYRPELLQEFQHQPEDFADLDEGLDSLTMLLSHSSGGGNMENKTTIP
DVFSDFDDDESSRSL*

>G1965 (1..609)

ATGGATAACTTCAATGTTGTTGCCAATGAAGACAATCAAGTGAATGATGTAAGCCTCCA
CCACCCCCACCGGAGTGTGTGCAAGATGTGATTCTGATAACACAAAATTTGTTACTAC

AACAATTATAGTGAGTTTCAACCGCGCTACTTCTGCAAGAACTGTGGAAGATACTGGACT
CATGGTGGGGCTTTAAGAAACGTACCAATTGGTGGGAGTAGTCGTGCAAGCGGACAAGG
ATAAATCAACCTTCAGTTGCTCAGATGGTTTCTGTTGGAATCCAACCGGGAACCGTTT
AGTTCTTTGTCTCATATTCATGGTGGTATGGTAACAAATGTGCATCCAACCTCAAACTTT
CGACCAAATCATCGCCTAGCTTTCCATAATGGATCATTTGAGCAAGATTATTATGATGTT
GGGTCTGATAATCTTTTGGTAAACCAACAAGTTGGTGGATATGTTGATAATCACAACGGT
TATCACATGAATCAAGTGGATCAATACAACCTGGAACCGAGCTTCAATAACGCTATGAAC
ATGAATTATAATAACGCTAGCACTAGCGGAAGGATGCATCCTAGTCATTTAGAGAAGGGT
GGTCCTTGA

>G1965 Amino Acid Sequence (domain in AA coordinates: 27-55)
MDNFNVANEDNQVNDVKPPPPPPRVCARCDSDNTKFCYNNYSEFQPRYFCNKCRRYWT
HGGALRNPVIGGSSRAKRTRINQPSVAQMVSVGIQPGNRFSSLSHIHGGMVTNVHPTQTF
RPNHRLAFHNGSFEQDYVDVGSNDLLVNQQVGGYVDNHNGYHNMVQDQYNWNQSFNNAMN
MNYNNASTSGRMHPSHLEKGGP*

>G1976 (1..1152)
ATGACTGATCCTTATTCCAATTTCTTCACAGACTGGTTCAAGTCTAATCCTTTTCACCAT
TACCCTAATTCTCCACTAACCCCTCTCCTCATCCTCTTCTCCTGTTACTCCTCCCTCT
TCCCTTCTTCTTCTTCCCTCAATCCGGAGACCTCCGCCGTCCACCGCCGCCACCAACTCCT
CCTCCTTCTCCTCTCTCCGAGAAGCCCTCCCTCTCCTCAGCCTCAGCCCCGCCAACAAA
CAACAAGACCACCATCACAACCATGACCACCTTATTCAAGAACCACCTTCAACCTCCATG
GATGTCGATACGATCATCATCAGATGATCATCATAACCTCGATGACGATGACCAT
GACGTCACCGTTGCTCTTCACATAGGCCTTCCAAGCCCTAGTGCTCAAGAGATGGCCTCT
TTGCTCATGATGTCTTCTTCTCCTCTTCTCCTCGAGGACCACTCATCATCAGAGGACATG
AATCACAAGAAAGACCTCGACCATGAGTACAGCCACGGAGCTGTGCGAGGAGGAGAAGAT
GACGATGAAGATTGATCGCGCGGAGACGGCGGCTGTAGAATCAGCAGACTCAACAAGGGT
CAATATTGGATCCCTACACCTTCTCAGATTCTCATTGGCCCTACTCAGTTCTCATGTCCT
GTTTGCTTCAAAACCTTCAACAGATACAATAACATGCAGATGCATATGTGGGACATGGA
TCACAATACAGAAAAGGACCTGAATCTCTAAGGGGAACACAACCAACAGGAATGCTAAGG
CTTCGGTGCTATTGCTCGCCCCAGGCTGTGCGCAACACATTGACCATCCAAGGGCAAAG
CCTCTCAAAGACTTTCAGAACCCTTCAAAACACATTACAAGAGAAAACATGGGATCAAACCT
TTCATGTGTAGGAAATGTGGAAAGGCTTTTCGAGTCCGAGGGGACTGGAGAACACATGAG
AAGAATTGTGGCAAACCTTTGGTATTGCATATGTGGATCTGATTTCAAGCACAAGAGATCT
CTCAAAGATCACATCAAGGCTTTTGGGAATGGTCATGGAGCCTACGGAATTGATGGGTTT
GATGAAGAAGATGAGCCTGCCTCTGAGGTAGAACAATTAGACAATGATCATGAGTCAATG
CAGTCTAAATAG

>G1976 Amino Acid Sequence (domain in AA coordinates: 219-323)
MTDPYSNFFTDWFKSNPFHHYPSNSTNPSPHPLPPVTPPSSFFFFPQSGDLRRPPPPPTP
PPSPPLREALPLLSLSPANKQQDHHNHHDHLIQEPPSTSMVDVYDHHHQDDHNLDDDDH
DVTVALHIGLPSPSAQEMASLLMSSSSSSSRTHHHEDMNHKKDLDEHYSHGAVGGED
DDEDSVGGDGGCRISRLNKGQYWIPTPSQILIGPTQFSCPVCFTFNRYNNMQMHMWGHG
SQYRKGPESLRGTQPTGMLRLPCYCCAPGCRNNIDHPRAKPLKDFRTLQTHYKRKHGIKP
FMCRCCKGKAFAVRGDWRTHKNCGLWYICGSDFKHKRSCLKDHIAFGNGHGAYGIDGF
DEEDEPASEVEQLDNDHESMQSK*

>G2057 (27..1289)
GCCGTCTCGACGAATAGTCTACCAATGTCTGACGACCAATTCCATCACCCGCCGCTC
CTTCTTCAATGAGGCACCGTTCTACGTCCGATGCGGCGGACGGCGGCTGCGGCGAGATTG
TTGAGGTGCAAGGTGGTCACATTGTTTCGGTCTACCGGAAGAAAAGACCGCCACAGCAAAG
TCTGCACGGCTAAAGGGCCACGTGACCGGCGGTGAGACTCTCTGCTCACACGGCGATT
AGTTTTACGATGTTCAAGACAGGCTTGGTTTCGACCGACCTAGCAAAGCCGTTGATTGGC
TTATCAAAAAGGCTAAGACTTCCATTGACGAGCTCGCTGAGCTTCTCCTCCCTGGAATCCCG
CCGATGCAATTGCGCTAGCCGCTGCTAACGCTAAACCCAGAAGAACCACGCCAAAACCC
AAATCTCTCCGTCTCCGCCACCGCCGCAACAGCAACAACAACAACAGCTTCAGTTTCG
GTGTTGGCTTCAACGGAGGAGGAGCAGAGCATCCGAGTAACAACGAGTCGAGTTTCTCC
CGCCGTCAATGGATTGAGATTGATAGCTGACACTATAAAGTCGTTTTTCCGGTGATTG
GCTCTTCAACGGAGGCTCCTTCGAATCATAACCTTATGCACAACATATCATCATCAGCATC
CGCCGGATTTGCTTTCTCGAACTAATAGCCAAAACCAAGATCTCCGTCTCTCGCTGCAAT
CGTTCCCGGATGGTCCACCGTCGCTTCTGCACCAACCAACATCACCACCAACACCTCTGCTT

CCGCCTCCGAGCCTACTCTGTTCTACGGACAGAGCAATCCGTTAGGGTTTGACACATCGA
GTTGGGAGCAGCAGTCGTCGGAATTCGGAAGGATTTCAGAGACTAGTGGCTTGAACAGCG
GCGGTGGCGGCGGAGCAACCGATACAGGAAACGGAGGAGGGTTTCTGTTTCGCTCCTCCTA
CTCCTTCAACGACGTCGTTTACGCCAGTTCTTGGCCAAAGCCAACAGCTTATTCTCAGA
GGGGTCCCTTTCAGTCCAGTTACAGTCCCAGTATCCGTGCTTGGTTTGATCCTCACCATC
ATCACCATCCCATCTCCACCAGATCTCAACCACCACCATCACCTTCTCCACCGGTTTC
ACCAATCAGCAATCCCCGGAATCGGATTCGCCTCAGGTGAATTCTCTTCGGGTTTTCGCA
TACCAGCACGGTTTCAGGGCCAAGAAGAGGAGCAGCACGACGGTCTCACTACAAGCCGT
CCTCTGCTTCTCTATTCTCGCCATTGACAATCGAACTAATCCTC

>G2057 Amino Acid Sequence (domain in AA coordinates: TBD)

MSDDQFHHPPPPSSMRHRSTSDAADGGCGEIVEVQGGHIVRSTGRKDRHSKVCTAKGPRD
RRVRLSAHTAIQFYDVQDRLGFDPRSKAVDWLIKAKTSIDELAEPLPPWNPADAIRLAAA
NAKPRRTTAKTQISPSPPPPQQQQQQQLQFGVGFNGGGAEHPSNNESSFLPPSMDSDSI
ADTIKSFPPVIGSSSTEAPSNHNLHMHNYHHQHPPDLLSRTNSQNQDLRLSLQSFDPGPPSL
LHHQHHTSASASEPTLFYQSNPLGFDTSWEQSSSEFGRIRLVWNSGGGGGATDT
GNGGGFLFAPPTPSTTSFQPVLGQSQQLYSQRGPLOSSYSPIRAWFDPHHHQSISTDD
LNHHHHLPPPVBHQAIPGIGFASGEFSSGFRIPARFQGEQEBEQHDGLTHKPSASSISRH
*

>G2107 (79..624)

ACCACAAAACAGAGCAACACACAACACAAAGCTTCATTTCAATTCTGTTTCGAGAACCTT
TTGAGAACCAGATCGGAGATGGAAAACGACGATATCACCGTGGCGGAGATGAAGCCAAAG
AAGCGTGCTGGACGGAGGATTTTCAAGGAGACACGTACCCAACTACAGAGGCGTGCGG
CGTAGGGACGGCGACAAATGGGTATGCGAAGTCCGTGAACCGATTATCAGCGTCGAGTC
TGGCTCGGAACCTATCCGACGGCAGATATGGCCGCACGTGCTCACGACGTGGCGGTCTT
GCTCTGCGCGGAGATCCGCGTGTGTTGAATTTCTCCGATTCTGCTTGGAGGTTGCCGGTG
CCGGCATCCACTGATCCGGACACGATCAGGCGCACGGCGGCCGAAGCAGCGGAGATGTTT
AGGCCGCCGGAGTTTAGTACAGGAATTACGGTTTACCCCTCAGCCAGTGAGTTTGACACG
TCGGATGAAGGAGTCGCTGGAATGATGATGAGGCTCGCGGAGGAGCCGTTGATGTCGCCG
CCAAGATCGTACATTGATATGAATACGAGTGTGTACGTGGACGAAGAAATGTGTTACGAA
GATTTGTCACTTTGGAGTTACTAAAATACGTATGTGTTAAAAAACCAAGATCGTATGTG
TATGTATGCATAATAAATGGGCTTAATGATGGGCATAGATATGATAGGTCCAGCCTATAT
GTTAAATGTGTTTTATTTTTTGGTTTATCTAGTTTCCTAGGTATTTACCAAATTGTATTA
GTATAAGTTTTATTAAAGAAATAATCAAAAATGTTGTTGCCAAAAAAAAAAAAAAAAAAAA
AAAAA

>G2107 Amino Acid Sequence (domain in AA coordinates: TBD)

MENDDITVAEMKPKKRAGRRIFKETRHPYIRGVRRRDGDKWVCEVREPIHQRRVWLGTYP
TADMAARAHDAVLALRGRSACLNFSDSAWRLPVPASTDPDTIRRTAAEAEMFRPPEFS
TGITVLPASAEFDTSDEGVAGMMRLAEPLMSPPRSVIDMNTSVYVDEEMCYEDLSLWS
Y*

>G211 (1..750)

ATGATGTCATGTGGTGGGAAGAAGCCAGTGTCTAAGAAAACAACGCCGTGTTGCACGAAG
ATGGGGATGAAGAGAGGACCATGGACGGTGGAGGAAGACGAGATTCTTGAGCTTCATT
AAGAAAGAAGGTGAAGGACGGTGGCGATCGCTTCTTAAGAGAGCTGGTTTACTCAGATGT
GGAAAGAGCTGTCGTCTACGGTGGATGAACTATCTCCGACCCTCGGTTAAACGTGGAGGA
ATTACGTCGGACGAGGAAGATCTCATCTCCGTCTTCCCGCCTCCTCGGCAACAGGTGG
TCATTGATCGCGGAAGGATACCGGAAGGACTGATAATGAAATTAAGAACTATTGGAAC
ACTCATCTTCGTAAGAAACTTTAAGGCAAGGAATTGATCCTCAAACCCACAAGCCTCTT
GATGCAAAACAACATCATAAACCAGAAGAAGAAGTTTCCGGTGGACAAAAGTACCCTCTA
GAGCCTATTTCTAGTTCTCATACTGATGATACCACTGTTAATGGCGGGGATGGAGATAGC
AAGAACAGTATCAATGTCTTTGGTGGTGAACACGGCTACGAAGACTTTGGTTTCTGCTAC
GACGACAAGTTCTCATCGTTTCTTAATTCGCTCATCAACGATGTTGGTGATCCTTTTGGT
AATATTATCCCAATATCTCAACCTTTCAGATGGATGATTGTAAGGATGGGATTGTTGGA
GCGTCGTCTTCTAGCTTAGGACATGACTAG

>G211 Amino Acid Sequence (conserved domain in AA coordinates:24-137)

MMSCGKKPVSKKTPCCTKMGMKRGPTVEEDEILVSFIKKEGEGRWRSIPKRAGLLRC
GKSCRLRWMNYLRPSVKRGITSDEEDLILRLHRLLGNRWSLIAGRIPGRTDNEIKNYWN
THLRKKLLRQGIDPQTHKPLDANNIHKPEEEVSGGQKYPLEPISSSHTDDTTVNGGDGDS

KNSINVFGGEHGYEDFGFCYDDKFSSFLNSLINDVGD PFGNIIPISQPLQMD DCKD GIVG
ASSSSLGHD*

>G2133 (26..457)

ATCTCATCTTCATCCACCCAAAAACATGGATTCAAGAGACACCGGAGAACTGACCAGAG
CAAGTACAAAGGTATCCGTCGTCGAAATGGGGAAAATGGGTATCAGAGATTCTGTGTCCT
GGGAACTCGTCAACGTCTCTGTTAGGCTCTTTCTCCACCGCAGAAGGCGTGCCGTAGC
CCACGACGTGCGTTTTTACTGCTTGACCGACCATCTTCCCTCGACGACGAATCTTTTAA
CTTCCCTCACTTACTTACAACCTCCCTCGCCTCCAATATATCTCCTAAGTCCATCCAAAA
AGCTGCTTCCGACGCCGGCATGGCCGTGGACGCCGATTCCATGGTGCTGTGTCTGGGAG
TGGTGGTTGTGAAGAGAGATCTTCCATGGCGAATATGGAGGAGGAGGACAACTTAGTAT
CTCCGTGTATGATTATCTTGAAGACGATCTCGTTTGATCTATACGAGTACGTTTTTAGCA
GTTAA

>G2133 Amino Acid Sequence (domain in AA coordinates:11-83)

MDSRDTGETDQSKYKGI RRRKWKVSEIRVPGTRQRLWLGSFSTAEGAAVAHDVAFYCL
HRPSSLDDSFNFP HLLTSLASNISPKSIQKAASDAGMAVDAGFHGAVSGSGGCEERSS
MANMEEEDKLSISVYDYLEDDLIV*

>G2134 (36..644)

GAGCAAAAACCTTTGTGTGCGTGTGTGTGTGTGTTTCATGGCTGGTCTTAGGAATTCGGTA
ACAGCGACAAAGCGCAAAACGATGGCAAAGGTGTACCATCTGCCTACAGAGGAGTCCGGA
AGAGAAAATGGGGGAAATGGGTGTCTGAAATCCGTGAACCGGGGACCAAGAACCGTATCT
GGCTAGGCAGTTTCGAGACTCCTGAAATGGCTGCAACCGCATAACGCTGGCAGCATTTC
ATTTTCAGAGGGAGAGAAGCTCGTCTCAACTTCCCTGAGCTCGCCAGCAGCCTTCCACGTC
CTGCAGACTCTAGCTCAGACAGCATTTCGATGGCAGTTCATGAGGCAACACTCTGCCGCA
CCACCGAAGGAACAGAGTCAGCCATGCAAGTGGACAGCTCAAGCTCCTCCAATGTAGCTC
CAACAATGGTCAGACTCTCGCCAGGGAAATTCAAGCGATCAACGAGTCAACTTTGGGAT
CTCCTACTACAATGATGCATTCAACATACGACCTTATGGAGTTTGTCTAATGATGTGGAGA
TGAATGCTTGGGAAACATACCAGAGTGACTTTCTTTGGGACCTTAACCCAAAACCTAA
CTCATGGAGAGCTTCTACAGCTCAATCTTACAATACCAGCATAAGTTACTGGCTTAGAAT
ACTTAAATTTATTGAAGTTTAGTTTTCAGAGTCTACCACAAGGGTTGTTGATTCTGACGT
TATAGCAAAGAATAAAGCTCATCAGATTTTGGAGGGAAAGACTCTATGAGCTTGATGGGT
CCCTGAAAGGACCTCTTCAAAATATTTTTAAATTTTTTGTACTAGTAGAAACATAGA
TTATGAGGTGTGACTTATTATTTATTTTTTACAATTGTTGTACCTCATTGATGTATTTG
ATTT

>G2134 Amino Acid Sequence (domain in AA coordinates: TBD)

MAGLRNSGNSDKAQNDGKGVPSAYRGVVRKRWKWKVSEIREPGTKNRIWLGSFETPEMAA
TAYDVAAFHFRGREARLNFP ELASSLPRPADSSSDSIRMAVHEATLCRTTEGTESAMQVD
SSSSSNVAPTMVRLSPREIQAINESTLGSPTTMMHSTYDPMEFANDVEMNAWETYQSDFL
WDP*PQNLTHGELLQLNLTIPIA*

>G2151 (236..1321)

TTTTTTTTTTAGGGTTCATAAGAACAAATTGGATTTTGTAGCTCACAGTATAAATAACCCG
ACTTTGATTACTGGGTAATTTTAAACCGCCATTGTTGTCTCTTTACTACTTTTGGGAA
TTAGGGTTTATGATTTCTGGGTATTAGATTAGATAAATTTGTTTCTTTTTTGTTAATC
AATTTAAAAATCTCTTATTTCTGTTAAAGACTTGTAATTTTGGAGTTTTTAATGCATGGA
CGGAAGAGAAGCAATGGCATTTCAGGCTCGCATTCTCAGTACTATCTTCAAAGAGGAGC
CTTTACTAATCTCGACCTTCCCAAGTCGCGAGTGGGCTTACGCGCCGCCGCCACATAC
GGGATTGAGGCCAATGTCTAACCTAACATTTCATCACCTTCAGGCTAACAAATCCAGGACC
TCCTTTCTCGGATTTTGGACACACCATTCACATGGGAGTGGTCTCCTCTGCTTCTGATGC
TGATGTGCAACCGCCACCGCCACCGCCACCAGAGGAACCGATGGTTAAGAGGAAACG
TGGACGGCCAAGAAAGTATGGAGAACCGATGGTTAGTAATAAGTCTAGGGACTCTTCTCC
AATGTCTGATCCTAATGAACCTAAACGGGCCAGAGGTCGACCTCCTGGAACCTGGAAGGAA
GCAACGCTTGGCTAATCTTGGTGAGTGGATGAATACTTCAGCTGGACTTGCTTTTGCACC
TCATGTGATCAGCATTGGAGCAGGAGAAGACATTGCTGCGAAAGTTTGTCAATTTTCA
ACAAAGACCTCGGGCTCTTTGTATAATGTGAGGCACTGGAACCATTTCTTCAGTCACTCT
GTGCAACCCCGGTTCAACCGATCGTCACTTAACATACGAGGGACCTTTTGTAGATTATAAG
TTTTGGTGGATCTTTTGTGGAATGAAGAAGGTGGATCCAGAAGTGAACAGGCGGATT
GAGTGTCTCTCTTCTCGTCCCGATGGTAGTATTATTGCCGTTGGAGTTGACATGCTTAT
CGCAGCCAACCTTGTTCAGGTGGTGGCATGTAGTTTTGTATACGAGCAAGGGCAAAGAC

TCATAATAACAATAACAAGACCATCAGACAAGAAAAGGAACCAAATGAAGAGGACAACAA
TAGTGAAAATGGAGACCACACCGGGTAGTGCAGCTGAACCAGCAGCATCTGCGGGTCAGCA
GACGCCACAGAACTTCTCTCTCAGGGAATAAGGGGGTGGCCCGGTTCAAGGCTCAGGCTC
TGGCAGATCACTTGACATTTGAGAAACCCACTCACTGATTTTGAATTTGACTCGTGGATG
ATATACACTATTAGTCTTTGAAGCAGCAGCATACAAAATGTGATTGCTGTACATATGTTA
TTGTAGATTTCTCTCTGCGGAATGTTGAAATCAGACATTTAAGGATTGATACTAGATCTCT
CAGCTCCTTCTAACATTGTTAATGTAACAGAACCCCTCCCACTTTTCATGCTATTTGC
>G2151 Amino Acid Sequence (domain in AA coordinates:93-113, 124-144)
MDGREAMAFPGSHSQYYLQRGAFNLAAPSQVASGLHAPPPHTGLRPMSPNPNHHPQANNP
GPPFSDFGHTIHMGVVSSASDADVQPPPPPPPEEPMVKRKRGRPRKYGEPMVSNKSRDS
SPMSDPNEPKRARGRPPTGRKQRLANLGEWMNTSAGLAFAPHVISIGAGEDIAAKVLSF
SQQRPRALCIMS GTGTISSVTLC KPGSTDRHLYEGPFEIISFGGSYLVNEEGGSRRTG
GLSVSLSRPDGSI IAGGVDMLIAANLVQVVACSFVYGARAKTHNNNNKTIRQEKEPNED
NNSEMETTPGSAAEPAASAGQQT P QNFSSQGI R WPGSGSGSGSRSLDICRNPLTDFDLTR
G*
>G2154 (82..1317)
GCAAAAAGAAAAATGAAAAAAATCCCTAACTCTCTCTCTCTAGAAATCTTATTTTTG
TGCGTATCTCTCTAAAAAGGAATGGATCCTAACGAAAGCCACCATCACCACCAACAACAA
CAGCTCCATCACCTCCACCAACAGCAACAGCAACAGCAGCAGCAACGACTCACTTCT
CCTTACTTCCACCACCAACTACAGCACCATCACCACCTTCCAACCACCGTAGCAACCACC
GCTTCTACCGGAAACGCCGTTCCATCTTCCAACAATGGGCTTTTCCCTCCGCAGCCTCAG
CCACAGCACCAAGCCTAATGATGGGTCTCTCTCTCGCGGTGTACCTCATTCACTTCCG
TCCTCGGTGTGACGGCGCGGATGGAGCCGGTAAAGAGGAAGAGGGGTGACCAAGAAAG
TATGTGACGCCGGAACAAGCCCTAGCGGCTAAGAAATGGCGTCTTCTGCGAGTAGTTCCG
TCTGCTAAACAGAGGCGAGAGCTTGCTGCTGTTACCGGTGGTACGGTATCGACTAATTCC
GGGTCAATCAAGAAATCTCAGCTTGGTTCTGTGCGGAAAACTGGACAATGTTTTACTCCG
CATATTGTTAATATAGCTCCTGGCGAGGATGTGGTCCAGAAAATTATGATGTTGCGAAAC
CAAAGCAAGCATGAAGTATGCGTTCTTCTGTCATCAGGCACTATCTCTAATGCATCCTTG
CGCCAACCGGCTCCATCAGGAGGCACTTACCATATGAGGGTCAATACGAGATTCTCTCA
CTATCTGGATCCTATATCCGAACCTGAACAAGGTGGTAAATCCGGCGGCCTTAGCGTTTCT
TTATCTGCTTCAGATGGTCAGATCATCGGTGGAGCGATTGGTAGCCATCTCACAGCTGCT
GGCCCGGTTCAAGGTGATTCTTGGTACGTTTCACTTGATAGAAAGAAGGATGCCGCCGGG
AGTGGTGGGAAAGGGGATGCTTCAAACAGTGGAAGTCGGTTAACTTCTCCTGTAAGCTCT
GGACAGTTGCTTGGCATGGGTTTCCCTCCTGGTATGGAATCTACGGGAAGAAATCCAATG
AGGGGAAACGACGAGCAACATGATCATCATCATCAAGCCGGTTTGGGTGGACCTCAT
CATTTTCATGATGCAAGCGCCGAGGGGATACACATGACACATTCCAGGCCATCTGAATGG
CGCGGAGGAGGCAACAGCGGTCTGATGGCAGAGCGGTGGCGGGTATGATTGTCAGGA
AGGATAGGACATGAGTCGTGCGGAGATGGAGATTACGAGCAGCAATACCGGATTAGCAG
AGCTTCCAGGAGAAAGTGTGTAGAGTTAGATCCCAAGTAGAGAAACAGAAGGCGAGCAAA
GAATCTGAAGTGAAGAGGACTTATTAGACAGAGACTCGTCTGAAGGGTCTTAAATCATA
GAAAGAAGTTGCTGAGTGATTGCTTTTGTCTTCTTCTTGGTACGGTGTATATATTAAC
TCCACAACCTTTTATATCTTTCAGTAACGATTCTCCTTCACTTTCAATTTTCATTCTCT
TTTTTTTATACTCTTTTCTTTTCTTATAATATTTTTTTTGGTTTTTCTTCTGTTTGT
CTAAAAAAGGAAATGCTCTTTTTGTGAAATATATACACTTCGTTTG
>G2154 Amino Acid Sequence (domain in AA coordinates:97-119)
MDPNESHHHHQQQLHHLHQQQQQQQQRLTSPYFHHQLQHHHHLPTTVATTASTGNAV
PSSNNGLFPPQPQPHQPNDSGLAVYPHVPSSAVTAPMEPVKRRGRPRKYVTPEQA
LAAKKLASSASSSSAKQRRELAAVTGTVSTNSGSSKSKSLGSGVKTGQCFTPHIVNIAP
GEDVVQKIMMFANQSKHELCLVLSASGTISNASLRQPAPSGGNLPYEQYIELSLSGSYIR
TEQGGKSGGLSVLSASDGQIIGGAIGSHLTAAGPVQVILGTFQLDRKKDAAGSGGKGDA
SNSGSRLTSPVSSGQLLGMGFPPGMESTGRNPMRGNDQHDHHLHQAGLGGPHHFMQAP
QGIHMTSRPSEWRGGGNSGHDGRGGGGYDLSGRIGHESSENGDYEQQIPD*
>G2157 (306..1238)
TCTTTTGAATTTAACCTTTTTTTCAGTAGCAAGCCAAAAAAGGACAAAGAGTT
CCTTTTATGATAAAGGTATGATGATAGCAACAAATGATACCCCATGTCTTGTGTGTCT
GCTTCATGCAACATGTTGGTTTGGATTGGTTAATCTAAAAGTTAAGATAAGGTTTTCG
GATTCTCTTCTGTCTTGTAAATAGTTTCTTGTGCGAGAGCCATCAACACCAACTTCAACA

AAAAAACAAGAAAAGAAAAGATTCTCTTTCTCGTTTTATTTCATTAGAGAAGAAAA
AAAGAATGGCGAATCCTTGGTGGGTAGGGAATGTTGCGATCGGTGGAGTTGAGAGTCCAG
TGACGTCATCAGCTCCTTCTTGCACCACAGAAACAGTAACAACAACACCCACCGACTA
TGACTCGTTCCGATCCAAGATTGGACCATGACTTCACCACCAACAACAGTGAAGCCCTA
ATACCCAGACTCAGAGCCAAGAAGAACAAGACAGACAGAGACGAGCAACCAGCTGTTGAAC
CCGGATCCGGATCCGGGTCTACGGGTGCTCGTCTAGAGGTAGACCTCCTGGTTCCAAGA
ACAAACCAAGAGTCCAGTTGTTGTACCAAAGAAAGCCCTAACTCTCTCCAGAGCCATG
TCTTGAGATTGCTACGGGAGCTGACGTGGCGGAAAGCTTAAACGCCCTTGCTCGTAGAC
GCGGCCGGGGCGTTTCGGTGTGAGCGGTAGTGGTTTGGTTACTAATGTTACTCTGCGTC
AGCCTGCTGCATCCGGTGGAGTTGTTAGTTTACGTGGTCAGTTTGAGATCTTGTCTATGT
GTGGGGCTTTTCTTCTACGTCTGGCTCTCCTGCTGCAGCCGCTGGTTTAACCATTTACT
TAGCTGGAGCTCAAGGTCAAGTTGTGGGAGGTGGAGTTGCTGGCCCGCTTATTGCCCTCTG
GACCCGTTATTGTGATAGCTGTACGTTTGTCAATGCCACTTATGAGAGGTTACCGATTG
AGGAAGAACAACAAGAGCAGCCGCTTCAACTAGAAGATGGGAAGAAGCAGAAAGAAG
AGAATGATGATAACGAGAGTGGGAATAACGGAAACGAAGGATCGATGCAGCCGCCGATGT
ATAATATGCCTCCTAATTTTATCCCAAATGGTCATCAAATGGCTCAACACGACGTGTATT
GGGGTGGTCTCCGCTCGTGTCTCCTTCGTATTGATTAGTTAGATAGGCGGTGGTTG
GTGCGTTCTTTTACTGGAATGATTATATTTTCCATTAGGATGGTTAGGCTTTTGTATT
TAAAGCTATCAAGTTTCTTTTTTTTTTACGGATAATTCGGATGACAATTAGCTAGTGT
GTTTGTGTTGTTTGTGGCGGCTTTCTGACTTGACTATTTTGTATCGCGGATAGCTTTGTA
TGAAAGTGAATTGATTGTAGAAATCGTCTTTTGAATTTTGTATGTTGGAACCA
>G2157 Amino Acid Sequence (domain in AA coordinates: 82-102, 164-107)
MANPWWVGNVAIGGVESPTSSAPSLHHRNSNNNPPTMTSRDPRLDHDFTTNNSGSPNT
QTQSQEEQNSRDEQPAVEPGSGSGSTGRRPRGRPPGSKNPKSPVVVTKESPNSLQSHVL
EIATGADVAESLNAFARRRGRVSVLSGSLVTNVTLRQPAASGGVSVLRGQFEILSMCG
AFLPTSGSPAAAAGLTIYLAGAQGVVGGVAGPLIASGPVIVIAATFCNATYERLPIEE
EQQQEQPLQLEDGKKQKEENDDNESGNNGNEGSMQPPMYNMPNFI PNHGQMAQHDVYWG
GPPPRAPPSY*
>G2181 (1..1005)
ATGATGCTTCGGTGAAGATGTGTTAAGCGAACTCGCCGGAGAAGAAAGGAACGAGAGA
GGATTGCCACCTGGCTTCCGGTTTCACCCGACGGACGAAGAGCTCATTACCTTCTACTTA
GCTTCCAAATCTTCCATGGTGGTCTCTCCGGCATTACATTTCCGAAGTTGATCTCAAC
CGCTGTGAACCTTGGGAGCTACCAGAAATGGCGAAGATGGGAGAGAGAGAGTGGTACTTT
TATAGTCTAAGGGACAGGAAATATCCGACAGGTTTGAGGACTAACAGAGCAACTACTGCT
GGATACTGGAAGCTACCGGCAAGATAAGGAAGTCTTCTCCGGCGGAGGAGGACAGCTT
GTTGGGATGAAGAAGAGTGTGGTGTCTACAAAGGTAGGGCTCCAGTGGCCTCAAGACT
AAGTGGGTGATGCATGAGTATCGCCTCGAAAACGACCATTACACCGCCACACGTGTAAG
GAGGAATGGGTGATTTGCAGAGTGTCAATAAAACAGGAGACAGAAAAATGTTGGATTA
ATCCATAACCAATCAGCTACCTTCATAACCATTCACTCTCAACAACACATCATCATCAT
CATGAAGCCTTACCTTTGCTTATAGAACCTTCCAACAAAACCTAACCAACTTCCCATCA
CTACTCTACGATGATCCACACCAAACTACAATAATAACAACCTTCTTCATGGATCATCA
GGCCACAACATCGACGAGCTCAAAGCCTTAATCAACCCGTGCTGTCTCTCAGCTCAACGGT
ATCATCTTTCCTTCAGGGAACAACAACAACGACGAAGACGACTTCGACTTTAACCTCGGC
GTGAAAACAGAGCAGTCTTCAACGGTAACGAAATTGACGTACGAGATTACTTGGAGAAC
CCTCTGTTTCAGGAAGCGAGTTATGGTCTGTTGGGTTTTTCGTCTTCTCCTGGACCTCTT
CACATGCTACTAGATTCTCCATGTCTTTAGGATTCCAGCTGTAG
>G2181 Amino Acid Sequence (conserved domain in AA coordinates:22-169)
MMLAVEDVLSELAGERNERGLPPGFRFHPTDEBLITFYLASKIFHGGLSGIHISEVDLN
RCEPWELPEMAKMGEREWFYFSLRDRKYPTGLRTNRATTAGYWKATGKDKEVFSGGGQL
VGMKKTLLVFYKGRAPRGLKTKWMHEYRLNDHSHRHTCKEEWVICRVFNKTGDRKNVGL
IHNQISYLNHNSLSTHHHHHEALPLLEPSNKTLTNFP SLLYDDPHQNYNNNNFLHGSS
GHNIDELKALINPVVSQLNGIIFPSGNMNNDEDDDFNLGVKTEQSSNGNEIDVRDYLEN
PLFQEAASYGLLGFS SPGLHMLLDSPCLGLFQL*
>G221 (115..795)
CTCTCTTATTCTCTCACTCTTTTTTTTTTATATTCTCTCTCTCTAAATCTATAAAATAT
ATTTAAAAACTTGATCGTATATAATAAAGTAAATAAAGAATAATAACAAAAAAATGGAG
AAAAGAGGAGGAGGAAGTAGTGGAGGTTCCGGATCATCAGCAGAAGCAGAAGTGAGAAAA

GGACCATGGACGATGGAAGAAGATCTTATTCTTATCAACTATATCGCCAACCACGGCGAT
GGTGTGTTGGAATTTCTCTCGCCAAATCTGCAGGTCTAAAACGAACCGGGAAAAGTTGCCGG
CTCCGGTGGCTGAACATATCTCCGCCCCGACGTACGACGGGGAAACATCACTCCAGAAGAG
CAACTTATCATCATGGAACCTTCATGCTAAGTGGGGAAAACAGGTGGTTCGAAAATCGCCAAA
CATCTTCCAGGAAGAACGGACAACGAGATCAAAAATTTCTGTAGGACAAGAATTCAAAAA
TACATCAAGCAATCGGATGTAACAACAACATCGTCCGTGGATCTCATCATAGCTCAGAG
ATCAACGATCAAGCTGCAAGCACGTGAGCCATAATGTCTTTTGTACACAAGATCAAGCG
ATGGAGACTTATTCTCCTACACCGACATCATATCAACATACCAATATGGAATTCAACTAT
GGTAACTATTTCGCCCGCGGCAGTGACGGCAACCGTGGATTATCCAGTACCGATGACCGTT
GATGATCAAACCGGTGAAAACATTTGGGGCATGGATGATATTTGGTCATCAATGCATTTA
TTGAATGGTAATTGATTGATCGGTGGACAAAACATGGAATATTAATTGAGTATTATATAT
GATTTTTAGGAGTACTATTATTAGTACGTGACATGTATATGTTTTTGCCTCGTTGTAGAG
GTTTGGGGTTTATAATATATATAATGTTATCTAATATGCAACCTTGATACATATTTGGA
TCTTTATTGAACCCATGTTATACATAAAATAAAATTTGTTGAAGGGTTCATAAAAAAAAAA
AAAAAAAAAAAAA

>G221 Amino Acid Sequence (domain in AA coordinates: 21-125)
MEKRGGGSSGSSGSAEVRKGPWTMEEDLILINYIANHGDGVWNSLAKSAGLKRTGKS
CRLRLWNLNRPDVRNRNITPEEQILIMELHAKWGNRWSKIAKHLPGRTDNEIKNFCRTRI
QKYIKQSDVTTTTSSVSGSHSSEINDQAASTSSHNVFCTQDQAMETYSPTPTSYQHTNMEF
NYGNYSAAAVTATVDYFVPMPTVDDQTGENYWGMDDIWSMHLNLGN*

>G2290 (119..982)

TTCTTTCTTTCTTTCTTTCTTCTTCTTCCAATCAAGAACAACCCCTAGCTCCTCTCTTTTCTC
TCTCTACCTCTCTTTCTCTATCTTCTCTTATCACTACTTCTCTCGCCGATCAATCATCAT
GAACGATCCTGATAATCCCGATCTGAGCAACGACGACTCTGCTTGGAGAGAACTCACACT
CACAGCTCAAGATTCTGACTTCTTCGACCGAGACACTTCCAATATCCTCTCTGACTTCGG
TTGGAACCTCCACCACTCCTCCGATCATCCTCACAGTCTCAGATTGACTCCGATTAAAC
ACAAACCACCGGAGTCAAACCTACCACCGTCACTTCTTCTTGTTCCTCATCCGCCGCCGT
TTCCGTTGCCGTTACCTCTACTAATAATAATCCCTCAGCTACCTCAAGTTCAAGTGAAGA
TCCGGCCGAGAACTCAACCGCCTCCGCCGAGAAAACACCACCACCGGAGACACCAGTGAA
GGAGAAGAAGAAGGCTCAAAAGCGAATTCGGCAACCAAGATTGCGATTTCATGACCAAGAG
TGATGTGGATAATCTTGAAGATGGATATCGATGGCGTAAATATGGACAAAAGCCGTCAA
GAATAGCCCCATTCCCAAGGAGCTACTATAGATGCACAAACAGCAGATGCACGGTGAAGAA
GAGAGTAGAACGTTTCATCAGATGATCCATCGATAGTGATCACAACATACGAAGGACAACA
TTGCCATCAAACATTGGATTCCCTCGTGGTGAATCCTCACTGCACACGACCCACATAG
CTTCACTTCTCATCATCATCTCCCTCCTCCATTACCAAATCCTTATTATTACCAAGAACT
CCTTCATCAACTTCACAGAGACAATAATGCTCCTTCACCGCGGTTACCCCGACCTACTAC
TGAAGATACACTTGCCGTGTCTACTCTCATCAGAGGAAGGCTTACTTGGTGATATTGTACC
TCAAATATGCGCAACCCCTTGAGGTAAGCTTGGTACGTAGCAATAGCTAAGGAGGTGCTA
ACTCATTATATATAGAAGATATTGCAGACCAGAATATGCGCAGGGAGGGTATAACAATAT
GGCGTTGTAACAATGGATCTATATATTACCTCATTGTTGATCAATAGCACACCACCGGTA
CGTTTGCAATTTCTTCATGTATATTTCTTGTATATATGTAGTTATATATCCAGGTATAA
TTTTGATGTAACACAACATTAATCTTAATCGTGGATCCATCCACATTTGATGCATGTAT
GTGCACTTAAGAAAAAGAACATGGAGGAAATAACGTTATTTTTATTATTCT

>G2290 Amino Acid Sequence (conserved domain in AA coordinates: 147-205)
MNDPDNPDLSNDDSAWRELTLTAQDSDFDRDTSNLSDFGWNLHHSSDHPHSLRFDSDL
TQTTGVKPTTVTSSCSSSAVSVAVTSTNNNPSATSSSSEDPAENSTASAETPPPETPV
KEKKKAQKRIRQPRFAFMKSDVDNLEDGYRWRKYQKAVKNSPFPRSYRCTNSRCTVK
KRVERSSDDPSIVFTTYEGQHCHQTIGFPRGGILTAHDPHSFTSHHHLPPPLPNPYQQE
LLHQLHRDNNAPSPRLPRPTTEDTPAVSTPSEGLLGDIVPQTMRNP*

>G2299 (231..941)

GCCAAAATTTTACCAACATTTTCTCTTCTCATATCAAAGTTTCTCTCTCATTTCTTCAT
CACACTTCACTGCCCTGTTTTTTTCTCTCATTTTGAATAGTTCTCAAACCTATATATTTT
TCCCCCTGAAGCCTAGCTATTTCTTTTATTTGCAATTAATCTCGGGATCCGAATCGAAAA
AAGCAATCAGAATAATAGACTTGTACGATACTTGTGCCTAAGCTAACACAATGGCAGAGG
AATACTACAGCCTCCGCTCGGAGAGAGTAACTCAGCTTCTTGTCCCTAACTCGGAGTCTG
ACTCAGTGAGTGACAAAAGCAAAGCTGAGCAAAGCGAGAAGAAGACTAAACGTGGGAGAG
ACTCCGGTAAACACCCTGTTTATCGCGGAGTAAGGATGAGGAACCTGGGGAAAATGGGTGT

CGGAGATTCTGAGCCGAGGAAGAAATCACGTATTTGGCTGGGAACTTTCCCGACGCCGG
AGATGGCGGCGCGTGCACACGACGTGGCGGCTCTGAGCATTAAAGGAACGGCCGCTATAC
TAAACTTCCCTGAACTCGCTGACTCATTCCCTCGACCGTTTCATTAAGCCCTCGAGACA
TTCAGACAGCAGCTCTTAAAGCAGCTCACATGGAACCGACGACGTGTTTTTCATCTTCCA
CGTCTTCGTCGTCGCTTTTGTCTTCTACGTCTTCGCTCGAGTCTCTTGTGTTGGTGATGG
ACCTCTCGAGGACTGAGTCGGAGGAGCTCGGTGAGATTGTGGAGCTTCCAAGTCTCGGGG
CGAGTTACGACGTGACTCGGCTAACCTTGGGAACGAGTTTGTCTTCTATGACTCAGTTG
ACTACTGTTTATATCCGCCGCCGTGGGGACAGTCGTCCGAAGATAACTATGGTCACGGAA
TTAGCCCTAATTTTGGCCATGGCTTGTCTATGGGATCTCTAACAGTTTATTTTGTATCATT
ACCATAATGTTTTGTTTAAACAGTTTATTTTGTATCATTGCCATAATGTTTTGTTTAAAT
CACGTTTTTAAACCCCTTTGCTGTTTTTGTTTTTTTTTTGTAGTTTTT

>G2299 Amino Acid Sequence (conserved domain in AA coordinates:48-115)

MAEYYSLRSEVRTQLLPNSES SVSDKSKAEQSEKTKRGRDSGKHPVYRGVVRMRNWG
KWVSEIREPRKKSRIWLGTFTPEMAARAHDVAALS IKGTAAILNFP ELADSFPRPVSL
PRDIQTAALKAAHMEPTTSFSSSTSSSSLSSTSSLES LVLVMDLSRTESEELGEIVELP
SLGASYDVDSANLGNEFVFYDSVDYCLYPPWPWGQSSSEDNYHGGISPNFGHLSWDL*

>G2340 (274..1275)

ATACAAAACCTCCCTCTTCTCTATCTTCTTCATCTTAAAGAAAAATAAGAGATATTCGTA
AAGAGAGAACACAAAATTTAGTTTACGAAAAGCTAGCAAAGTCGAGTATCGAGGAATAA
CAGAATAAGACGTATCTATCCTTGCCTTAATGTTCTTACCAAAGATCTAGTCCCTTCTT
TGTATGATCGATCCATCACAAGCCCAACAACAACAACTACATCTCTTCTCTATCTCT
AGCTTCTATTTTTTAATACATTCAAGAATCAAGAATGGTACGGACGCCGTGTTGTAGAGCA
GAAGGGTTGAAGAAAGGAGCATGGACTCAAGAAGAAGACCAAAAGCTTATCGCCTATGTT
CAACGACATGGTGAAGGCGGTTGGCGAACCCCTTCCGGACAAAGCTGGACTCAAAGATGT
GGCAAAAGCTGCAGATTGAGATGGGCGAATTACTTAAGACCTGACATTAAACGTGGAGAG
TTTAGCCAAGACGAGGAAGATTCCATCATCAACCTCCACGCCATTGATGGCAACAAATGG
TCGGCCATAGCTCGTAAAAATACCAAGAAGAACAGACAATGAGATCAAGAACCATTGGAAC
ACTCACATCAAGAAATGTCTGGTCAAGAAAGGTATTGATCCGTTGACCCACAAATCCCTT
CTCGATGGAGCCGGTAAATCATCTGACCATTTCCGCGCATCCCGAGAAAAGCAGCGTTTCT
GACGACAAAGATGATCAGAATTCAAATAACAAAAAGTTGTCAGGATCATCATCAGCTCGG
TTTTTGAACAGAGTAGCAAACAGATTCCGTCATAGAATCAACCACAATGTTCTGTCTGAT
ATTATTGGAAGTAATGGCCTACTTACTAGTCACACTACTCCAACATAAGTGTTCAGAA
GGTGAGAGGTCAACGAGTTCTTCTCCACACATACCTCTTCAATCTCCCATCAACCGT
AGCATAACCGTTGATGCAACATCTCTATCCTCATCCAGTTCTCTGACTCCCCCGACCCG
TGTTTATACGAGGAAATAGTCGGTGACATTGAAGATATGACGAGATTTTCATCAAGATGT
TTGAGTCATGTTTTATCTCATGAAGATTTATTGATGTCCGTTGAGTCTTGTTTGGAGAAT
ACTTCATTCATGAGGGAAATTACAATGATCTTTCAAGAGGATAAAATCGAGACGACGTCG
TTTAATGATAGCTACGTGACGCCGATCAATGAAGTTGATGACTCCTGTGAAGGGATTGAC
AATTATTTTGGATGAGTTATATTGATGATGATGAAAATTTGCATTTGGCATGTAAATCAA
TTAGAGTTTGATTGCTATGGTGTTTTGTAGTTTGTGTGTAGTGTGTTTCGACCGTCAA
AAAAAAAAAAAAAAAAAAAAAAAAAAAAA

>G2340 Amino Acid Sequence (domain in AA coordinates:14-120)

MVRTPCCRAEGLKKGAWTQEEDQKLIAYVQRHGEQWRTLPDKAGLKRCGKSCRLRWANY
LRPDIKRGEFSQDEEDSIINLHAIHGNKWSAIARKIPRTDNEIKNHWNTHIKKCLVKKG
IDPLTHKSLLDGAGKSSDHSAPKSSVHDDKDDQNSNNKKLSGSSSARFLNRVANRFGH
RINHNVLSDIIGSNGLLTSHTTPTTSVSEGERSTSSSSTHTSSNLPINRSITVDATSLSS
STFSDSPDPCLYEIVGDIEDMTRFSSRCLSHVLSHEDLLMSVESCLENTSFMREITMIF
QEDKIETTSFNDSYVTPINEVDDSCGIDNYFG*

>G2346 (1..1011)

ATGGAGTTGTTAATGTGTTCCGGGTCAGGCCGAGTCAGGTGGTCTTCTTCCACCGAGTCT
TCTTCACTCAGTGGTGGACTCAGGTTTGGTCAGAAGATCTACTTCGAGGATGGATCCGGA
TCCAGAAGCAAGAACCGGGTCAATACCGTTCGTAAGTCGTCTACCACGGCGAGGTGCCAA
GTGGAAGGTTGTAGAATGGATCTAAGCAATGTTAAAGCTTATTACTCGAGACACAAAGTT
TGTTGCATTCACTCTAAATCATCTAAAGTCATTGTCTCTGGTCTTCATCAAAGGTTTTGT
CAACAATGTAGCAGGTTTACCAGCTTTCTGAGTTTGACTTGAGAGAAAAGAAGTTGTGCG
AGAAGACTCGCTTGTCTATAACGAACGACGAAGAAAACCACAACCCACAACGGCTCTTTTC
ACTTCTCATTACTCTCGAATCGCTCCATCTCTTTACGGAAACCCCAATGCTGCAATGATT

AAAAGCGTTTTGGGAGATCCTACTGCGTGGTCAACCGCAAGATCAGTGATGCAGCGGCCT
GGACCGTGGCAGATTAATCCAGTTAGGGAAACCCATCCACACATGAATGTTTTATCACAT
GGAAGCTCAAGCTTTACTACATGTCCAGAGATGATAAACAACAATAGCACAGATTCAAGC
TGTGCTCTCTCTCTCTGTCAAACCTCATACCCAATTCATCAGCAGCAACTTCAGACACCA
ACAAATACATGGCGACCATCTTCTGGTTTCGACTCGATGATCTCATTCTCCGATAAGGTT
ACAATGGCTCAGCCACCGCCCATTTCAACCCATCAGCCGCCCATCTCAACACATCAGCAG
TACCTCAGCCAAACTTGGGAAGTCATCGCGGGCGAAAAGAGCAATTCACATTATATGTCT
CCTGTGAGTCAAATCTCGAGCCAGCAGATTTCCAGATAAGCAATGGCAGTGTGTGCCCC
TATTCTCCTCCGTCCTTACTATCTCTTGTGTGCTACTTGC GGCCGCTATAG
>G2346 Amino Acid Sequence (domain in AA coordinates: 59-135)
MELLMSGQAESGSSSTESSLSGGLRFGQKIYFEDGSGSRSKNRVNTVRKSSSTARCQ
VEGCRMDLSNVKAYYSRHKVCCIHSKSKVIVSGLHQRFCQQCSRFBHQLSEFDLEKRSCR
RRLACHNERRRKQPPTALFTSHYSRIAPSLYGNPNAAMIKSVLGDPTAWSTARSVMQRP
GPWQINPVRETHPHMNVLSHGSSSFTTCPEMINNNSTDSSCALSLLSNSYPIHQQLQTP
TNTWRPSSGFDMSISFDKVTMAQPPPISHTQPPISTHQYLSQTWEVIAGEKSN SHYMS
PVSQISEPADFQISNGSVSPSPPSLLSLVLCYLRPL*
>G237 (1..852)
ATGGCGAAGACGAAATATGGAGAGAGACATAGGAAAGGGTTATGGTCACCTGAAGAAGAC
GAGAAGCTAAGGAGCTTCATCCTCTCTTATGGCCATTCTTGTCTGGACCACCTGTTCCCATC
AAAGCTGGGTTACAAAGGAATGGGAAGAGCTGCAGATTAAGATGGATTAATTACCTAAGA
CCAGGGTTAAAGAGGGATATGATTAGTGCAGAAGAAGAAGAGACTATCTTGACGTTTCAT
TCTCCCTTGGGTAACAAGTGGTCGCAAATAGCTAAATCTTACC GGGAAGAACAGACAAT
GAGATAAAGA ACTATTGGCACTCTCATTGAAAAAGAAATGGCTCAAGTCTCAGAGCTTA
CAAGATGCAAAATCTATTTCCCTCCTTCGTCTTCATCATCATCACTTGTGTGCTGTGGA
GAAAGAAATCCGGAACCTTGATCTCGAATCACGTGTTCTCCCTCCAGAGACTTCTAGAG
AACAAATCTTCATCTCCCTCACAAGAAAGCAACGGAAATAACAGCCATCAATGTTCTTCT
GCTCCTGAGATTCCAAGGCTTTTCTTCTCTGAATGGCTTTCTTCTTCATATCCCCACACC
GATTATTCCTCTGAGTTTACCGACTCTAAGCACAGTCAAGCTCCAAATGTGCAAGAGACT
CTCTCAGCTTATGAAGAAATGGGTGATGTTGATCAGTTCATTACAACGAAATGATGATC
AACAAACAGCAACTGACTCTTAAACGACATGTGTTTGGTTCCAAATGTAAGAAGCAGGAG
CATCATATTTATAGAGAGGCTTCAGATTGTAATTCTTCTGCTGAATTCTTTTCTCCACCA
ACAACGACGTAAATTGCGTTTATTGTAATGTAAATCAAATTTCTAAGGCAAAACCGGAAA
AAAAAAAAAAAAAAAAAAAA
>G237 Amino Acid Sequence (domain in AA coordinates: 11-113)
MAKTKYGERHRKGLWSPEEDEKLRSFILSYGHSCWTTVPKAGLQRNGKSCRLRWINYLR
PGLKRDMSIAEEETILTFHSP LGNKWSQIAKFLPGRTDNEIKNYWHS LKKKWLKSQL
QDAKSISPPSSSSSLVACGERNPETLISNHVFS LQRLLENKSSSPSQESNGNNSHQCSS
APEIPRLFFSEWLSSSYPHTDYSEFTDSKHSQAPNVEETLSAYEEMGDVDQFHYNEMMI
NNSNWT LNDIVFGSKCKKQEHHIYREASDCNSSAEFFSPPTTT*
>G2373 (48..1199)
GCAAAATCCTCAGATCGTCTTACCTTCTCCGAATCGATCGATTTTTTCATGGAGGACGACG
ACGAGATTCAATCAATTCATCTCCGGGAGATTCTTCCCTTTCACCACAAGCTCCTCCTT
CTCCGCCGATTTTGCCAACAAACGACGTGACGGTGGCCGTCGTGAAGAAACCACAACCGG
GGCTTTCTCTCAATCTCCGTCCATGAACGCTTTAGCGTTAGTGGTTCATACTCCTTCTG
TAACCGGTGGTGGTAGCGGAAACAGAAACGGACGAGGAGGAGGAGGAGGAAGCGGTG
GTGGTGGAGGAGGAGAGATGATTGTTGGAGCGAAGAAGCTACAAAGGTTCTAATCGAAG
CTTGGGGAGATCGATTCTCTGAACCAGGTAAAGGAAC TTTGAAGCAACAACATTGGAAG
AAGTAGCTGAGATTGTGAACAAGAGTCGTCAATGCAAATACCTTAAAAC TGATATTCAGT
GTAAGAACAGAATTGATACGGTGAAGAAGAAGTATAAGCAAGAGAAAGCTAAGATTGCTT
CTGGTGATGGACCTAGTAAATGGGTTTCTTCAAGAAGCTTGAGAGTTTGATTGGTGGTA
CTACAACATTCATTGCTTCTTCAAAGCTTCAGAGAAGGCTCCTATGGGAGGAGCTCTTG
GGAATAGCCGTTTCGAGTATGTTTAAACGGCAAACTAAAGGTAATCAGATTGTGCAGCAAC
AACAAGAGAAGAGAGGCTCTGATTGATGCGGTGGCATT TTAGGAAACGTAGTGCTTCTG
AGACTAGTCTGATCTGATCCTGAACCTGAGGCTTCTCTGAGGAATCTGCTGAGAGTC
TCCCACCTTTGCAACCGATTCAACCGCTTTCGTTT CATATGCCAAAGCGGTTGAAGGTGG
ATAAGAGTGGAGGTGGAGGGAGTGGAGTTGGAGATGTGGCGAGGGCGATACTTGGATTTA
CGGAAGCTTATGAGAAGCGGAAACTGCTAAGCTTAAGTTAATGGCGAACTGGAAAAGG

AGAGGATGAAATTTGCTAAAGAGATGGAGTTGCAGAGAATGCAGTTCTTGAAAACCTCAAT
TGGAGATAACACAGAACAATCAAGAAGAGGAAGAGAGGAGCAGGCAGCGAGGAGAAAGGA
GGATCGTTGATGATGATGATGATCGCAATGGCAAGAATAACGGCAATGTAAGTAGCTGAC
AATTGAACACACAAATGTTCTATGATATTTGCTATGATAAGCTGGATTTTAGGTTTGA
TGG

>G2373 Amino Acid Sequence (domain in AA coordinates:290-350)
MEDDDEIQSIPSPGSSLPQAPPSPILPTNDVTVAVVKPQPGLSSQSPSMNALALVV
HTPSVTGGGSGNRNGRGGGGSGGGGGGRDDCWSEETKVLIEAWGDRFSEPGKGTLLKQ
QHWKEVAEIVNKSQRCKYPKTDIQCKNRIDTVKKYKQEKAKIASGDGPSKWVFFKKLES
LIGGTTTFIASSKASEKAPMGALNSRSSMFKRQTKGNQIVQQQEKRGSDSMRWHFRK
RSASETESESDPEPEASPEESAESLPPLQPIQPLSFHMPKRLKVDKSGGGSGVGDVARA
ILGFTEAYEKAETAKLKLMAELEKERMKFEMELQRMQFLKTQLEITQNNQEEEEERSRQ
RGERRIVDDDDDRNGKNNGNVSS*

>G2376 (39..1370)
CACGAGCTTCTGACTCAGATCCGGCGATATCGAATTCCATGGAGGACGATGAAGACATCC
GATCTCAGGGTTCCGATTCACCTGATCCGTCTTCTCTCCCGCCGGCGGGACGAATCACGG
TTACGGTGGCTTCGGCAGGTCCGCCTTCTTATCTCTGACTCCTCCGGGTAATTCGTGCG
AGAAGGATCCGGATGCGTTGGCTCTGGCGCTGCTTCCGATTACAGGCCAGCGGTGGAGGGA
ATAACAGCAGTGGGAGACCAACCGGCGGCGGGAGGGAGGATTGTTGGAGCGAAGCAG
CTACGGCTGTGTTGATTGATGCGTGGGGTGAGAGATACTTGGAGCTTAGCAGAGGGAATC
TGAAGCAGAAGCACTGGAAAGAGGTGGCTGAGATTGTGAGCAGCAGAGAGGATTACGGTA
AAATTTCCAAAACCTGATATACAGTGTAAAGAAATAGGATCGATACGGTGAAGAAGAAGTATA
AACAAGAGAAGGTGAGAATCGCTAACGGCGGTGGCCGTAGCAGATGGGTGTTCTTCGACA
AGCTTGACCGTCTGATTGGATCAACGGCGAAGATCCCGACGGCAACTTCTGGAGTCAGCG
GTCTGTGCGGAGGATTGCATAAGATTCTATGGGTATTCCAATGGGAAGTCGTTTGAATC
TGTACCATCAGCAAGCTAAGGCTGCAACACCGCTTTCAATAATCTTGACCGGTTAATTG
GAGCTACGGCTAGAGTCTCAGCTGCTTCTTTCGGTGGCAGTGGTGGAGGAGGCGGAGGAG
GATCTGTCAATGTACCTATGGGAATTCGATGAGTAGCCGTTTCACTCCGTTTGGACAGC
AAGGGAGGACTCTGCCACAGCAAGGTAGGACACTGCCACAGCAACAGCAGCAAGGGATGA
TGGTGAAGAAGGTGTAGTGAGTCAAAACGCTGGCGTTTTCAGGAAGAGGAACGCTTCTGATT
CAGACTCGGAATCTGAAGCAGCAATGTGATGATTCCGGTGACAGTTTACCACCTCCTC
CTCTGTGCAAGAGGATGAAGACGGAGGAGAAGAAGAAGCAAGATGGTGTGAGTGGGGA
ACAAATGAGGGAGCTGACTCGGGCAATCATGAGATTCCGGTGAAGCTTATGAGCAAACAG
AGAATGCGAAACTGCAACAGGTGGTTGAGATGGAGAAAGAGAGGATGAAGTTCTTGAAGG
AGCTTGAGTTGCAGAGAATGCAGTTCTTTGTGAAGACTCAATTGGAGATATCAAACTTA
AGCAGCAACATGGGAGGAGAATGGGAAACACCAGTAATGATCATCATCAGCCGCAAGA
ACAACATCAATGCGATTGTCAACAACAACAACGATTGGGTAATAACTAGAATTTAGTGA
TGCAGTGTCTGAATTGATATATTTTAGATTTGAG

>G2376 Amino Acid Sequence (domain in AA coordinates:79-178, 336-408)
MEDDEDIRSQSDSPDPSSPPAGRITVTVASAGPPSYSLTPGNSSQKDPDALALALLP
IQASGGNNSSSRPTGGGGREDWCSEAATAVLIDAWGERYLELSRGNLQKQHWKEVAEIV
SSREDYGIKPKTDIQCKNRIDTVKKYKQEKVRIANGGGRSRWVFFDKLDRLLIGSTAKIP
TATSGVSGPVGGLHKIPMGIPMGSRSNLYHQQAATPPFNNLDRLLIGATARVSAASFEG
SGGGGGGGSVNVPMGIPMSRSAPFGQQGRTLPPQQGRTLPPQQQQGMMVKRCSESKRWR
RKRNASDSDSEEAAMSDDSGDSLPPPLSKRMKTEEEKKKQDGDGVGNKWRELTRAIMRF
GEAYEQTENAKLQQVVEMEKERMKFLKELELQRMQFFVKTQLEISQLKQQHGRRMGNTSN
DHHHSRKNNINAIVNNNNDLGNN*

>G24 (194..724)
CGGACGCGTGGGCAAATATTAAAATAAAAAGTGTGCGGTGAATTCTCAATCTTTGTCTTCT
TTCGTCTCTCTTTAAACTCTCTCCGTCCCTCCTTATTATGTAACCGTCTCGCCGTCAA
TTTTCAAAATCTCTCCCTCCGTTCATAAACCCAGATCGAAATTTATGGTTTTGTAAATTT
TTTACCGCGGTTATGGAGACGAAGCGGCGGTGACAGCGACGGTTACGGCGGCGACGAT
GGGATTTGGGACGAGGAAGAGAGATCTGAAACCGTATAAAGGAATACGAATGAGGAAATG
GGGGAATGGGTGGCGGAGATACGGGAACCGAATAAGAGATCAAGGATCTGGTTAGGTTCT
TTATGCGACGCTGAAGCGGCGGCGAGAGCTTACGACACTGCTGTTTTTTACCTCCGTGG
TCCTTCAGCGAGGCTTAATTTTCCGAGCTTTTGGCTGGACTTACTGTTTCTAACGGCGG
AGGAAGAGGTGGTGATTTATCGGCGGCGTATATTAGGAGAAAAGCGGCGGAGGTTGGTGC

TCAGGTTGATGCGCTTGGAGCGACGGTGGTTGTGAATACCGCGCGGAGAATCGCGGTGA
TTACGAGAAGATTGAGAATTGTCGTAAGAGCGGTAACGGGTCATTGGAACGGGTCGATTT
GAATAAATTACCCGACCCGAAAATTCGGATGGTGATGATGACGAATGTGTGAAAAGAAG
ATAGAAAAAATAAAAAGTAGTTGTAGAAGGAGAGACGAGAATGTTGTCTTTAAGATGCG
CTGTTGCCGCTAACATGCGCTTTTCGATTTTAGTGTTAAACATGCGCCTCCATTGTTTTTG
GGTTTTGTTTTTCGTCGTCGATAATCAAAGATTTTAAACACAATTCTCAAATTTTTCACT
TGTTACAAACTAGATTTGCATGATCTTTGTATTAAACGAATAACGATTAAGTCCTAAA

>G24 Amino Acid Sequence (domain in AA coordinates: 25-93)

METEAAVTATVTAATMGIGTRKRDLPYKGIIRMRKWKWVAEIREPNKRSRIWLGSYATP
EAAARAYDTAVFYLRGPSARLNFPELLAGLTVSNGGGRGDLAAYIRRKAAEVGAQVDA
LGATVVVNTGGENRGDYEKIENCRKSGNGSLERVDLNLKLPDPENSDDDDCEVKRR*

>G2424 (1..999)

ATGAGGATGGAGATGGTGCATGCTGACGTGGCGTCTCTCTCCATAACACCTTGCTTCCCG
TCTTCTTTGTCTTCGTCCTCACATCATCACTATAACCAACAACAACATTGTATCATGTCTG
GAAGATCAACACCATTTCGATGGATCAGACCACTTCATCGGACTACTTCTCTTTAAATATC
GACAATGCTCAACATCTCCGTAGCTACTACACAAGTCATAGAGAAGAAGACATGAACCTT
AATCTAAGTGATTACAGTAATTGCAACAAGAAAGACACAACAGTCTATAGAAGCTGTGGA
CACTCGTCAAAAGCTTCGGTGTCTAGAGGACATTGGAGACCAGCTGAAGATACTAAGCTC
AAGAAGTACTGCGCGTCTACGGTCCACAAAACCTGGAACCTCATAGCTGAGAAGCTCCAA
GGAGATCCGGGAAAAGCTGTAGGCTTCGATGGTTTAACCAACTAGACCCAAGGATAAAT
AGAAGAGCCTTCACTGAGGAAGAAGAAGAGAGGCTAATGCAAGCTCATAGGCTTTATGGT
AACAAATGGGCGATGATAGCGAGGCTTTTCCCTGGTAGGACTGATAATTCTGTGAAGAAC
CATTGGCATGTTATAATGGCTCGCAAGTTTAGGGAACAATCTTCTTCTTACCGTAGGAGG
AAGACGATGGTTTTCTCTTAAGCCACTCATTAAACCCTAATCCTCACATTTCAATGATTTT
GACCCTACCCGGTTAGCTTTGACCCACCTTGCTAGTAGTGACCATAAGCAGCTTATGTTA
CCAGTTCCTTGCTTCCCAGGTTATGATCATGAAAATGAGAGTCCATTAATGGTGGATATG
TTCGAAACCCAAATGATGGTTGGCGATTACATTGCATGGACACAAGAGGCAACTACATTC
GATTTCTTAAACCAAACCCGGAAGAGTGAGATATTTGAAAGAATCAATGAGGAGAAGAAA
CCACCATTTTTTCGATTTTCTTGGGTTGGGGACGGTGTGA

>G2424 Amino Acid Sequence (conserved domain in AA coordinates:107-219)

MRMEMVHADVASLSITPCFPSSLSSSSHHYNNQQHCIMSEDQHHSMQDQTTSSDYFSLNI
DNAQHRLRSYYTSHREEDMNPNSDYSNCKNDTIVYRSCGHSSKASVSRGHWPAEDTKL
KELVAVYGPQNWNLIAEKLQGRSGKSCRLRWFNQLDPRINRRAPTEEEEEERLMQAHRLYG
NKMAMIARLFPGRTDNSVKNHWHVIMARKFREQSSSYRRRKTMSVLKPLINPNPHIFNDF
DPTRLALTHLASSDHKQLMLPVPCFPGYDHENESPLMVMDFETQMMVGDYIAWTQEATTF
DFLNQTGKSEIFERINEEKKPPFFDFLGLGTV*

>G2505 (1..1026)

ATGGGTTCTTCGTCGAACGGAGGAGTGCCACCTGGTTTTCCGGTTTCATCCGACGGACGAA
GAGCTTCTCCATTACTACTTGAAGAAGAAAATCTCTTACCAAAAGTTTGAGATGGAAGTC
ATCAGAGAGGTTGACTTAACAAGCTTGAGCCTTGGGATTTGCAAGAGAGATGCAAGATA
GGATCAACACCACAAAACGAATGGTACTTCTTCAGCCACAAGGACAGGAAATATCCGACG
GGGTCAAGGACCAACCGTGCTACTCATGCAGGGTTCTGGAAGGCGACGGGACGTGACAAG
TGCATAAGGAACCTTACAAAAAGATAGGAATGAGGAAGACACTTGTGTTCTACAAAGGT
AGAGCTCCTCATGGCCAAAAGACTGATTGGATCATGCATGAGTACCGTCTTGAAGACGCT
GATGATCCTCAAGCCAACCTAGTGAAGATGGATGGGTGGTATGTAGAGTGTATGAAG
AAAAATTTGTTCAAGGTAGTAAATGAAGGTAGCTCAAGCATTAACTCATTGGACCAACAC
AACCATGACGCATCTAACAACAACCATGCACCTCAAGCTCGTAGCTTTATGCACCGAGAC
AGTCCATACCAGCTAGTACGTAACCACGGAGCCATGACATTGCAACTTAACAAGCCTGAC
CTTGCTCTTCATCAATACCCACCAATCTTCCACAAGCCACCTTCACTTGGATTTGACTAC
TCTTCAGGACTTGCAAGGGACAGTGAGAGTGCGGCTAGTGAAGGGTTACAATACCAGCAA
GCGTGTGAGCCGGGTTTAGACGTTGGTACATGTGAGACAGTGGCTAGTCATAATCATCAA
CAAGGTCTAGGTGAATGGGCAATGATGGATAGACTTGTGACTTGTACATGGGAAATGAA
GATTCCTCTAGAGGGATTACGTATGAGGATGGTAACAACAATTCGTCCTCTGTGGTTCAG
CCAGTTCCTCGCGACGAACCAGCTAACATTGCGTAGTGAGATGGATTTCTGGGGTTATTCT
AAATAG

>G2505 Amino Acid Sequence (domain in AA coordinates: 10-159)

MGSSSNNGVPPGFRFHPTDEELLHYLLKKKISYQKFEMEVIREVLDLNLKLEPWLQERCKI

GSTPQNEWYFFSHKDRKYPTGSRTNRATHAGFWKATGRDKCIRNSYKKIGMRKTLVIFYKG
RAPHGQKTDWIMHEYRLDADDPQANPSEDGWVVCVFMKKNLFKVNEGSSSINSLDQH
NHDASNNHALQARSFMHRDSPYQLVRNHGAMTFELNKPDLALHQYPPIFHKPPSLGFDY
SSGLARDESAASEGLQYQQACEPGLDVGTCETVASHNHQQGLGEWAMMDRLVTCHMGNE
DSSRGITYEDGNMNSSSVVQVPATNQLTLRSEMDFWGYSK*

>G2512 (64..798)

AACTTAGTGCCACTTAGACACAATAAGAAAACCGTTAACAAGAAGAAAAAAGATCG
AAAATGGAATATCAAACTAAGTTCTTAAGTGGAGAGTTTCCCCGGAGAACTCTTCTTCA
AGCTCATGGAGCTCACAAGAATCATTCTTGTGGGAAGAGAGTTTCTTACATCAATCATTT
GACCAATCCTTCCTTTTATCTAGCCCTACTGATAACTACTGTGATGACTTCTTTGCATTT
GAATCATCAATCATAAAAGAAGGAAGCCACCGTGGCGGCCGAGGAGGAG
AAGTCATACAGAGGAGTGAGGAAACGGCCGTGGGGGAAATTGCGGCCGAGATAAGAGAC
TCAACGAGGAAAGGATAAGAGTGTGGCTTGGGACATTGACACCGCGGAGGCGGCGGT
CTCGCTTATGATCAGGCGGCTTTCGCTTTGAAAGGCAGCCTCGCAGTACTCAATTTCCCC
GCGGATGTCGTTGAAGAATCTCTCCGAAGATGGAGAATGTGAATCTCAATGATGGAGAG
TCTCCGGTGATAGCCTTGAAGAGAAAACACTCCATGAGAAACCGTCTAGAGGAAAGAAG
AAATCTTCTTCTTCTTCGACGTTGACATCTTCTCCTTCTTCTCCTCCTCCTATTCTCT
TCTTCGTCCTTCTTCTTCTTGTGCTCAAGAAGTAGAAAACAGAGTGTGTTATGACGAA
GAAAGTAATACACTTGTGTTCTTGGAGATTAGGTGCTGAATACTTAGAAGAGCTT
ATGAGATCATGTTCTTGATAATCTGCTTCTACAAATTTTATGTAATTGA

>G2512 Amino Acid Sequence (conserved domain in AA coordinates: 79-139)

MEYQTNFLSGEFSPENSSSSSSWSSQESFLWEESFLHQSFQSFLLSSPTDNYCDDFFAFE
SSIIKEEGKEATVAEEEEESYRGVVRKRPWGKFAAEIRDSTRKGIRVWLGTFTDAEAAAL
AYDQAAFALKGSLAVLNFPADVVEESLRKMENVNLNDGESPVIALKRKHSRMRNRPRGKKK
SSSSSTLTSSPSSSSSYSSSSSSSSSLSSRSRKQSVVMTQESNTTLVVLEDLGAELYLEELM
RSCS*

>G2513 (69..698)

TTTCAACAGTAATTTAAGTTAACCGGAGTCTCTTTTGTGTTTCCGGCGAATTTTGGTAC
TTTGAGTTATGAATAATGATGATATTATTCTGGCGGAGATGAGGCCAAGAAGCGTGCGG
GAAGGAGAGTGTTTAAGGAGACAGTCACCCAGTTTACAGAGGCATAAGGCGGAGGAACG
GTGACAAATGGGTCTGCGAAGTCAGAGAACCGACGCACCAACGCCGATTTGGCTCGGGA
CTTATCCACAGCAGATATGGCAGCGCGTGCACACGACGTGGCGGTTTGTAGCTCTGCGTG
GGAGATCCGCATGTTTGAATTTCCGCCGACTCCGCTTGGCGGCTTCCGGTGCCGGAATCCA
ATGATCCGGATGTGATAAGAAGAGTTGCGGCGGAAGCTGCGGAGATGTTTAGGCCGGTG
ATTTAGAAAGTGAATTACGGTTTTCCTTGTGCGGAGATGATGTGGATTGCGGTTTGTG
GTTCCGGTTCGGCTCTCGTTTCGGGATCGGAGGAGAGGAATTTCTTCTCGTATGGATTG
GAGACTACGAAGAAGTCTCAACGACGATGATGAGACTCGCGGAGGGGCCACTAATGTGCG
CGCCGCGATCGTATATGGAAGACATGACTCCTACTAATGTTTACACGGAAGAAGAGATGT
GTTATGAAGATATGTCAATGTGGAGTTACAGATATTAAGTGGGACTCACATATCTACTAT
ACATAATATTTAGCTTTTATGTAAGAGGTATTTATGTGAGTTTAAAGATTGTAGATGTGT
CCCAGGCGTTAGAAGTTTCCTTGATGTTATGGAATCTTTGTACCTATAAAATTATAAAAT
T

>G2513 Amino Acid Sequence (domain in AA coordinates: TBD)

MNNDIILAEMRPKKRAGRRVFKETRHPVYRGIRRRNGDKWVCEVREPTHQRRIWLGTYT
TADMAARAHDAVAVLALRGRSACLNFADSAWRLPVPESNDPDVIRRVAAEAEMFRPVDLE
SGITVLPACAGDDVDLGFSGSGSGSGSEERNSSSYGFGDYEEVSTTMMRLAEGPLMSPPR
SYMEDMTPTNVYTEEMCYEDMSLWSYRY*

>G2519 (83..694)

CAAAGTGAAAACATAAGATCATCTTCTTCGTTGATAGATCAATATAGGAAGTCCAGAAGA
GAATCTTGATCAATTAAGTATCATGTCTCACATCGCTGTTGAAAGGAATCGAAGAAGGCA
AATGAACGAGCATCTTAAATCCCTTCGTTCTTTGACTCCTTGTCTTACATCAAAAGGGG
AGATCAAGCTTCGATCATCGGAGGAGTGATAGAGTTCATCAAAGAGTTGCAGCAATTGGT
TCAAGTTCTTGAGTCCAAGAAACGTCGAAAGACCTAAACCGACCATCTTCCCTTATGA
TCACCAGACAATCGAGCCATCCAGTTTAGGAGCCGCCACTACCCGAGTACCGTTTAGTTCG
AATCGAAAATGTGATGACCACAAGTACTTTCAAGGAAGTAGGAGCATGCTGTAACCTCCCC
TCATGCTAACGTAGAAGCAAAGATTTAGGTTCTAATGTTGTATTGAGAGTTGTCTCTAG
GCGAATCGTGGGGCAGCTCGTAAAGATCATCTCTGTCTTAGAGAAGCTATCTTTCAAGT

TCTTCACCTCAATATTAGTAGCATGGAGGAGACTGTCTTATACTTTTTCGTTGTTAAGAT
AGGATTGGAGTGTCACTTAAGCTTGGAGGAGCTAACTCTTGAAGTTCAGAAAAGCTTTGT
GTCTGATGAAGTGATCGTCTTACCAATTAAAAACAAAATTCTACATGTACTAGAGCGTG
TATCGTTTTTTGGGATTAATAATCATATAATCGTTACATGAGCCTTGATACTTTGCTAGA
AATAAGCTCCTCTAAACAAAACCTTCTTTTTAAAAAAACACACTTATGTTTTACTTAGTT
TGTTGTTGTATCCGAAGTTGATCAACGTTGTAATTTCCACAATAAATCATGACATTTTA
TATGCTCT

>G2519 Amino Acid Sequence (domain in AA coordinates:1-65)

MSHIAVERNRRRRQMNEHLKSLRSLTPCFYIKRGDQASIIGGVIEFIKELQQLVQVLESKK
RRKTLNRPSFPYDHQTIIEPSSLGAATTRVPSRIENVMTTSTFKEVGACCNSPHANVEAK
ISGSNVVLRVVSRRIVGQLVKIISVLEKLSFQVLHLNISSMEETVLYFFVVKIGLECHLS
LEELTLEVQKSFVSDEVIVSTN*

>G2520 (133..1197)

AAGGAGTTTTGCATACTCACCAAGCCACAATCATTTCTCTCTCTATCTCTCTGGTTT
TGAATCGGCGACGACTGAGTCAACTCGGTGTTGTTACTGGTTTCGTCGTATGTGTTGTAA
CTGATTAAGTTGATGGATCCGAGTGGGATGATGAACGAAGGAGGACCGTTAATCTAGCG
GAGATCTGGCAGTTTTCCCTTGAACGGAGTTTCAACCGCCGGAGATTCTTCTAGAAGAAGC
TTCGTTGGACCGAATCAGTTTCGGTGATGCTGATCTAACCACAGCTGCTAACGGTGATCCA
GCGCGTATGAGTCACGCGTTGTCTCAGGCGGTTATTGAAGGTATCTCCGGCGCTTGAAAA
CGGAGGGAAGATGAGTCTAAGTCGGCGAAGATCGTCTCCACCATTGGCGCTAGTGAAGGT
GAGAACAAAAGACAGAAGATAGATGAAGTGTGTGATGGGAAAGCAGAAGCAGAATCGCTA
GGAACAGAGACGGAACAAAAGAAGCAACAGATGGAACCAACGAAAGATTATATTCATGTT
CGAGCTAGAAGAGGTCAAGCTACTGATAGTCACAGTTTAGCTGAAAGAGCGAGAAGAGAG
AAAATAAGTGAGCGGATGAAAATCTTGCAAGATCTTGTTCCGGGATGTAACAAGGTTATT
GGAAAAGCACTTGTTCTAGATGAGATAATTAACTATATACAATCATTGCAACGTCAGATT
GAGTTCTTATCGATGAAGCTTGAAGCAGTCAACTCAAGAATGAACCTTGGTATCGAGGTT
TTTCCACCCAAAGAGGTGATGATTCTCATGATCATCAACTCAATCTTCTCCATTTTTTTC
ACAAAACAATACATGTTTCTATCGAGGTATTCTCGGGGTAGGAGTCTCGATGTTTATGCG
GTTCCGTCATTTAAGCATTTGCAATAAACGGAGTGACCTCTGTTTTTGTCTGCTCCCA
AAAACAGAACTTAAGACAACTATATTTTCAAAAACATGACATGTTTCTGTGATATTCT
CGAGTAGGAGTTCGCTATTAGTTTCATCTAAGCATTGCAATGAACCGTTTGGTCAGCAAGCG
TTTGAGAATCCGGAGATACAGTTCGGGTGCGAGTCTACGAGGGAATACAGTAGAGGAGCA
TCACCAGAGTGGTTGCACATGCAGATAGGATCAGGTGGTTTCGAAAGAACGTCCTGA

>G2520 Amino Acid Sequence (domain in AA coordinates: 135-206)

MDPSGMNNEGSPFNLAIEWQFPLNGVSTAGDSSRRSFVGPNQFGDADLTAAANGDPARMS
HALSQAVIEGISGAWKRREDESKAKIVSTIGASEGENKRQKIDEVCDGKAESAELGTET
EQKKQOMEPTKDYIHVRARRGQATDSHSLAERARRREKISERMKILQDLVPGCNKVIKAL
VLDEIINYIQLRQVEFLSMKLEAVNSRMNPGIEVFPPKEVMILMIINSIFSIFFTKQY
MFLSRYSRGRSLDVYAVRSFKHCNKRSDLCFCSCSPKTELKTTIFSQNMTCFCRYSRVGV
AIISSKHCNEPFGQAFENPEIQFGSQSTREYSRGASPEWLHMQIGSGGFERTS*

>G2533 (1..1080)

ATGATAAGCAAGGATCCAATATCGAGTTTACCTCCAGGGTTTCGATTTTCATCCAACAGAT
GAAGAACTCATTCTCCATTACCTAAGGAAGAAAGTTTCTCTTCCCCAGTCCCGCTTTCG
ATTATCGCCGATGTGATATCTACAAATCCGATCCATGGGATTTACCAGCTAAGGCTCCA
TTTGGGGAGAAAGAGTGGTATTTTTTTCAGTCCGAGGGATAGGAAATATCCAAACGGAGCA
AGACCAACAGAGCAGCTGCGTCTGGATATTGGAAAGCAACCGGAACAGATAAATTGATT
GCGGTACCAAATGGTGAAGGGTTTCATGAAAACATTGGGTATAAAAAAGCTCTTGTTGTTT
TATAGAGGAAAGCGTCCAAAAGGTGTTAAAACCAATTGGATCATGCATGAATATCGTCTT
GCCGATTCAATTATCTCCAAAAGAAATTAAGTCTTCTAGGAGCGGTGGTAGCGAAGTTAAT
AATAATTTTGGAGATAGGAATTCTAAAGAAATATTCGATGAGACTGGATGATTGGGTCTT
TGCCGGATTATACAAGAAATCACACGCTTCATTGTCTATCACCTGATGTTGCTTTGGTCACA
AGCAATCAAGAGCATGAGGAAAATGACAACGAACCATTCGTAGACCGCGGAACCTTTTGG
CCAAATTTGCAAAATGATCAACCCCTTAAACGCCAGAAGTCTTCTGTTGTTCTCAAAAC
TTACTAGACGCTACAGATTTGACGTTTCTCGCAAATTTTCTAAACGAAACCCCGGAAAAT
CGTCTGAATCAGATTTTCTTTTCATGATTGGCAATTTCTCTAATCCTGACATTTACGGA
AACCATTACTTGGATCAGAAGTTACCGCAGTTGAGCTCTCCACTTCAGAGACAAGCGGC
ATCGGAAGCAAAAGAGAGAGAGTGGATTTTGGCGGAAGAAACGATAAACGCTTCGAAGAAG

ATGATGAACACATATAGTTACAATAATAGTATAGATCAAATGGATCATAGTATGATGCAA
CAACCTAGTTTCTGTAACCGAACTCATGATGAGTTCTCACCTTCAATATCAAGGCTAG
>G2533 Amino Acid Sequence (conserved domain in AA coordinates:11-186)
MISKDPISSLPPGFRFHPTEELILHYLRKKVSSSPVPLSIIADVDIYKSDPWLPAKAP
FGEKEWYFFSPDRKYPNGARPNRAASGYWKATGTDKLIAPNNGEGFHENIGIKKALVF
YRGKPPKGVKTNWIMHEYRLADSLSPKRINSSRSGGSEVNNNFGRNSKEYSMRLDDWVL
CRIYKKS HASLSSPDVALVTSNQEHENDNEPFVDRGTFLPNLQNDQPLKRQKSSCSFSN
LLDATDLTFLANFLNETPENRSBSEDFSMIGNFSNPDIYGNHYLDQKLPQLSSPTSETSG
IGSKRERVDAEETINASKMMNTYSYNNSIDQMDHSMMQQPSFLNQELMMSSHLQYQG*

>G2534 (1..975)

ATGGATAATATAATGCAATCGTCAATGCCACCGGGATTCCGATTTTCATCCGACAGAGGAA
GAGCTTGTGGGTTATTACCTAGATAGGAAGATCAATTCAATGAAGAGTGCTTTAGATGTC
ATTGTAGAGATTGATCTCTACAAAATGGAGCCATGGGATATACAAGCGAGGTGTAAACTA
GGGTATGAAGAGCAAAACGAGTGGTACTTCTTTAGTCATAAGGACAGGAAGTACCCTACC
GGGACTAGGACCAACCGAGCCACTGCGGCTGGGTTCTGGAAAGCCACGGGTAGAGACAAG
GCGTACTATCAAAAACAGTGTCTCGGAATGCGGAAGACACTTGTCTACTACAAGGGT
CGAGCTCCTAATGGAAGAAAGTCCGATTGGATCATGCACGAATACCGTCTCCAAAACCTCC
GAGCTTGCCCCGGTTTCAGGAGGAAGGCTGGGTGGTGTGTGCGAGCATTTAGGAAGCCAATT
CCAAACGAGGCCATTAGGGTACGAGCCATGGCAGAACCAGCTCTACCACGTCGAAAAGT
AGTAACAACACTACTCATCTTCAGTGACAATGAACACGAGTCATCATATCGGTGCATCTTCA
TCAAGTCATAACCTTAATCAAATGCTCATGAGCAATAACCACTACAATCCTAATAATACA
TCCTCATCGATGCATCAATATGGCAACATTGAGCTCCCGCAGTTGGACAGCCCGAGCTTG
TCGCCTAGTTTGGGACGAATAAAGATCAGAACGAGAGTTTCGAGCAAGAAGAAGAGAAG
AGCTTTAACTGTGTGGATTGGAGAACACTAGATACCTTGCTTGAGACACAAGTCATACAT
CCGCATAACCCTAATATTCTTATGTTTCAAACGAGTCGTATAATCCGGCGCCAAGCTTC
CCTTCCATGCATCAAAGCTATAATGAGGTGCAAGCTAATATTCATCATTTCTTGGATGC
TTCCCTGACTCGTAA

>G2534 Amino Acid Sequence (conserved domain in AA coordinates:10-157)

MDNIMQSSMPPGFRFHPTEELVGYLDRKINSMKSLADVIVEIDLYKMEPWDIQARCKL
GYEEQNEWYFFSHKDRKYPTGTRTNRATAAGFWKATGRDKAVLSKNSVIGMRKTLVYYKG
RAPNGRKSDWIMHEYRLQNSELPVQEEGWVVCRAFRKPIPNQRPLGYEPWQNQLYHVES
SNYSSSVTMTNTSHHIGASSSSHNLNQMLMSNNHYNPNNTSSSMHQYGNIELPQLDSPSL
SPSLGNTKDQNESFEQEEKSFNCVDWRTLDLTLETQVIHPHNPILMFETQSYNPAPSF
PSMHQSYNEVEANIHHS LGCFPDS*

>G2573 (34..957)

CCAGATTTAATTTGAGACTCTCAAAGAAACACCATGGAAGAAGAGCAACCTCCGGCCAAG
AAACGAAACATGGGGAGATCTAGAAAAGGTTGCATGAAAGGTAAGGCGGTCCAGAGAAC
GCCACGTGTACTTTCCGTGGAGTTAGGCAACGGACTTGGGGTAAATGGGTGGCTGAGATC
CGTGAGCCTAACCGTGGGACTCGTCTCTGGCTCGGCACGTTTAATACCTCGGTGAGGCC
GCCATGGCTTACGATGAAGCCGCTAAGAAACTCTATGGACACGAGGCTAACTCAACTTG
GTGCACCCACAACAACAACAAGTAGTAGTGAACAGAACTTGTCTTTTTCTGGCCAC
GGGTGCGGTTCTTGGGCTTATAATAAGAAGCTCGATATGGTTCATGGGTTGGACCTTGGT
CTCGGCCAGGCAAGTTGTTACGAGGTTCTTGCTCAGAGAGATCGAGTTTCTACAAGAA
GATGATGATCATAGTCATAATCGATGTTTCGTCTTCAAGTGGTTTCAATCTTTGTTGGTTA
TTACCTAAACAAAGTGATTACAAGATCAAGAGACCGTTAATGCTACGACTAGTTATGGC
GGTGAAGGCGGTGGTGGCTCTACGTTAACGTTTTTCGACCAATTTGAAACCAAGAATTTG
ATGAGTCAGAAATTATGGATTATACAATGGAGCTTGGTCTAGGTTTCTTGTGGGGCAAGAA
AAGAAGACGGAACATGACGTGTCTCGTGTGGATCGTCGGACAACAAGGAGAGTATG
TTGGTTCCTAGTTGCGGCGGAGAGAGGATGCATAGGCCGGAGTTGGAAGAGCGAACAGGA
TATTTGGAATGGATGATCTTTTGGAGATTGATGATTTAGGTTTGTGATTGGCAAAAAT
GGAGATTTCAAGAATTGGTGTGTGAAGAGTTTCAACATCCATGGAATTGGTTCTGAGAG
TTTTTATTTATTACTATTATTATCATACATATTTCTTATTTGACTTAGG

>G2573 Amino Acid Sequence (domain in AA coordinates: TBD)

MEEEEQPPAKKRNMGSRKGMKGKGGPENATCTFRGVRQRTWGKWAIEIREPNRGRRLWL
GTFNTSVEAAMAYDEAAKLYGHEAKLNLVHPQQQQQVVNRLNLSFGHSGSGSWAYNKKL
DMVHGLDLGLGQASCSRSRGSFSLQEDDDHSHNRCSSSSGSNLCWLLPKQSDSQDQE
TVNATTSYGGEGGGSTLTFTSNLKPKNLMSQNYGLYNGAWSRFLVGQEKTEHDVSSSC

GSSDNKESMLVPSGGERMHRPELEERTGYLEMDDLLEIDDLGLLIGKNGDFKNWCCEE
FQHPWNWF*

>G2589 (23..1354)

AAAGAAAAGAAAAATAAGATAATGAGGACGAAGACTAAGTTAGTACTCATACCTGATAG
ACACTTTTCGGAGAGCCACATTCAGGAAGAGGAATGCAGGGATAAGGAAGAACTCCACGA
GCTGACAACTCTCTGTGACATCAAAGCATGTGCGGTAATCTACAGTCCGTTGAGAATCC
AACGGTGTGGCCGTCACCGAAGGTGTTCAAGAGGTGATTTTCGGAGTTCATGGAGAAGCC
GGCGACAGAACGGTCCAAGACGATGATGAGTCATGAGACTTTCTTGCGGGACCAAATCAC
CAAAGAACAAAACAACTAGAGAGTCTACGTCGTGAAAACCGAGAACTCAGCTTAAGCA
TTTTATGTTTGATTGCGTTGGAGGCAAGATGAGTGAGCAACAGTATGGTGCAAGGGACCT
TCAAGATTTTAAGTCTTTTACTGATCAATATCTTAATCAGCTTAATGCCAGGAAGAAGTT
CCTTACAGAATATGGTGAGTCTTCTTCTCTGTTTCTCTCTCTGTTTGATGTTGCGGGTGC
CAATCCTCCTGTTGTTGTCAGATCAAGCTGCGGTAAGTCTCTCTCTTGTGTTGCTGTTGC
GGGTGCCAATCTTCTGTTGTTGCTGATCAAGCTGCGGTAAGTCTCTCTCTGTTTGC
TGTGCGGGTGCCAACTTCTCTGTTGTTGTCAGATCAAGCTGCGGTTAATGTTTCTACTGG
ATTTTCATAACATGAATGTGAACAGAAATCAGTATGAGCCGGTTTCAGCCCTATGTCCCTAC
TGGTTTTAGTGATCATATTCAATATCAGAATATGAACCTCAATCAAAACCAACAAGAGCC
GGTTCATTACAGGCTCTTGCTGTTGCGGGTGCCGGTCTTCTATGACTCAGAATCAGTA
TGAGCCCGTTCACTACAGAGCTTTGCTGTGCGGGTGCGGTTCTTCTATGAGTCAGTT
GCAGTATGAGCCGGTTTCAGCCTTATATCCCTACTGTTTTTAGTGATAATGTTCAATATCA
GCATATGAATTTGTATCAAAATCAACAAGAGCCGGTTCACTACCAAGCTCTTGGTGTGTC
AGGTGCCGGTCTTCTTATGAATCAGAATCAGTATGAGCCGGTTTCAGCCCTATGTCCCTAC
TGGTTTTAGTGATCATTTTCAGTTTGAGAATATGAATTTGAATCAAAATCAACAGGAGCC
GGTTCAATACCAAGCTCCTGTTGATTTTAATCATCAGATTCACAAAGGAACTATGATAT
GAATTTGAACCAGAATATGAGTTTGGATCCAAATCAGTATCCGTTTCAAAATGATCCATT
CATGAATATGTTGACAGAATATCCTTATGAATAAGCGGGTTATGTTGGAGAGCATGCAC

>G2589 Amino Acid Sequence (domain in AA coordinates: TBD)

MRTKTKLVLPDRHFRRATFRKRNAGIRKKLHELTTLCDIKACAVIYSPFENPTVWPSTE
GVQEVISEFMKPAERSKTMMSHETFLRDQITKEQNKLES LRRENRETQLKHFMDCVG
GKMSEQQYGARDLQDLSLFTDQYLNQLNARKKFLTEYGESSSSVPPLFDVAGANPPVVAD
QAAVTVPPLFAVAGANLPVVADQAAVTVPPLFAVAGANLPVVADQAAVNVPFGFHNMNVN
QNQYEPVQPYVPTGFSDDHIQYQNMNFQNNQEPVHYQALAVAGAGLPMTQNQYEPVHYQS
LAVAGGGLPMSQLQYEPVQPYIPTVFSDDNVQYQHMLYQNNQEPVHYQALGVAGAGLPMT
QNQYEPVQPYVPTGFSDDHFFQFENMNLNQNQEPVQYQAPVDFNHQIQQGNYDMNLNQNMS
LDPNQYFPQNDPFMNMLTEYPYE*

>G2687 (45..1139)

CTCTGTCTCTCGTATCTTTCTACTACTCTGTTTCTTGAATTCTAATGAACAACATCGACG
ACGCAAAGACGGAGACTTCAGTGTCTTCAGGTTCAAGCGACTCTTTCTTGCTCTCAAGA
AACGCATGAGACTTGATGACGAACAGAAAACGCCCTAGTGGTTTCGTCTTCACCAAAGA
CGGTTGTGGCTTCTGGCAATGTCAAGTACAAAGGAGTCGTTTCAGCAACAGAACGGTCATT
GGGGTGCCAGATTTACGCAGACCACAAAAGGATTTGGCTTGGAACCTTTCAAATCCGCTG
ATGAAGCCGCCACGGCTTACGATAGTGCATCTATCAAACCTCCGAAGCTTTGACGCTAACT
CGACCCGGAACCTTCCCTTGCTTACAATCACTCTCAACGAACAGACTTTCAAATTTGCT
ACACAACAGAGACTGTGTTGAACATGATCAGAGACGGTTCGTACCAACACAAATTCAGAG
ATTTTCTCAGAATCAGATCTCAGATTGTTGCGAGTATCAACATCGGGGGACCAAAACAAG
CCCGAGGAGAAGTGAATCAAGAATCAGACAAGTGTCTTTCTTGACACAGCTTTTTCAGA
AGGAATTGACACCGAGCGATGTAGGGAACTAAATAGGCTTGTGATACCTAAAAAGTATG
CAGTGAAGTATATGECTTTTCATAAGCGCTGATCAAAGCGAGAAAGAAGAGGGTGAAATAG
TAGGATCTGTGGAAGATGTGGAGGTTGTGTTTTACGACAGAGCAATGAGACAATGGAAGT
TTAGGTATTGTTACTGGAAGTAGCCAGAGCTTTGTCTTCACCAGAGGATGGAATAGTT
TCGTGAAGGAGAGAATCTCAAGGAGAAGGATGTTATTGCCTTCTACACTTGCGATGTCC
CGAACAAATGTGAAGACATTAGAAAGGTCAAAGAAAAGAACTTCTTGATGATCGATGTTT
GCTTTTCAGACAACGGTTCGGTGTAGCTGAGGAAGTAAGTATGACGGTTCATGACAGTT
CAGTGCAAGTAAAGAAAACAGAAAACCTTGGTTAGCTCCATGTTAGAAGATAAAGAAACCA
AATCAGAGGAGAACAAAGGAGGTTTATGCTGTTTGGTGAAGGATCGAATGTCCTTAGG
GAATTTTCTTTAAAAGTTTCTTACTTCAACTAGAACTTGTTTTACTTGTACCT

>G2687 Amino Acid Sequence (domain in AA coordinates: TBD)

MNNIDDAKTETSVSSGSSDSFLPLKKRMRLLDEPENALVVSSSPKTVVASGNVVKYKGVVQ
QQNGHWGAQIYADHKRIWLGTFKSADEAATAYDSASIKLRSFDANSHRNFPWSTITLNEP
DFQNCYTTETVLNMIRDGSYQHKFRDFLRIRSQIVASINIGGPKQARGEVNQESDKCFSC
TQLFQKELTPSDVGKLNRLVIPKKYAVKYMFPISADQSEKEEGEIVGSVEDVEVVFYDRA
MRQWKFRYCYWKSSQSFVFTRGWNSFVKEKNLKEKDVIAFYTCVDPNNVKTLEGQRKNFL
MIDVHCFSDNGSVVAEBVSMTVHDSSVQVKKTENLVSSMLEDKETKSEENKGGFMLFGVR
IECP*

>G27 (83..622)

CAAAATACCAAAAACAAAACATTTTTTTTAATCTTCCCACCAATTTTTTCTCTTTCTCT
CGTTACATTAAATTATCTTTAGATGCAAGACTCTTCCCTCTCAGCAATCGCAACGTAACCT
CCGGTCACCGGTGCCGGAGAAAACCGAAAGAGTTCTAAGACTAAAAATGAGCAAAAAGG
TGTTTCTAAACAACCAAAATTTTCGTGGGGTCAGAATGAGACAATGGGGAAAATGGGTGTC
TGAAATTAGAGAACCAAGAAAGAAATCAAGAATATGGCTCGGTACTTTCTCTACGCCGGA
GATGGCGGCGCGTGACACGACGTGGCGGCTTTAGCCATCAAAGGTGGCTCTGCCACCT
TAATTTCCCGGAGCTAGCTTACCATTGTCCGAGACCGGTAGCGCGGACCCTAAAGACAT
TCAAGAAGCCGCCGCCGACGAGCTGCCGTTGACTGGAAGCACCGGAGTCTCCGTCTAG
CACCGTGACGTCACTCCAGTCGCCGACGACGCTTTCTCCGATCTTCTGTATCTTTTGCT
TGACGTGAATGATCAACAACAAAACGATGAGATTCTGGGACTCGTTTCCGTACGAAGATCC
TTTTCTTCTTGGAAAAATTACTAGAAGGCAAAATCTTGCCGGCGAACGGATTTTCCGGTGGT
TTCCCGGTAAATAAGAAGACGATGTCGTTTTGTACCTTTTTTGTCTACGATGGGAAATTT
CTTTTTTTTTTACGTGTGAGTAAAAGTTTCCGAATGTGTGATGTGTAAGTAAGTACAGGT
TATTTAAATTTCTTTTTTTTTGTACAAATACGTACGTACATTACCAAAAAGTTTTCATTTATT
GTGCTTTTATCTTCCAAATTCATTAATAAAAAAAAAAAAAAAAAA

>G27 Amino Acid Sequence (domain in AA coordinates: 37-104)

MQDSSSHESQRNLRSPVPEKTGKSSKTKNEQKGVSKQPNFRGVRMRQWGWVSEIREPRK
KSRIWLGTFSPTEMAARAHDVAALAIKGGSAHLNFPPELAYHLPRPASADPKDIQEAAAAA
AAVDWKAPESPSSSTVTSSPVADDAFSDLPDLLLDVNDHNKNDGFWDSPFYEDPFFLENY*

>G2720 (1..894)

ATGGAAGCGAAGAAGGAAGAGATAAAGAAAGGTCCATGGAAAGCCGAAGAAGACGAAGTA
CTCATCAACCATGTCAAGAGATACGGTCTCTCGTGATTGGAGCTCCATTCGATCCAAAGGT
CTTCTTCAACGCACCGGCAAAATCCTGTCTGCTTTCGTGGGTCAATAAACTCCGTCCCAAT
CTCAAAAATGGATGCAAGTTCTCGGCTGACGAAGAGAGGACTGTGATTGAGTTACAATCT
GAGTTTGGTAACAAATGGGCGAGAATCGCTACGTATCTACCGGAAGAACTGATAACGAT
GTGAAGAATTTCTGGAGTAGCAGACAAAAGAGACTCGCTAGGATTCTTCATAACTCCTCT
GATGCATCGAGTTTCGAGTTTCAATCCCAATCTTCTTCTCATCGACTCAAGGGCAAA
AACGTCAAACCAATCCGTCAATCCTCTCAGGGTTTTGGTTTTGGTTGAGGAAGAGGTTACA
GTTTCTTCTTCATGTTCCAGATGGTTCTTATTCATCTGATCAAGTTGGTGATGAAGTC
TTGAGGTTGCCGATTTGGGTGTTAAGTTAGAGCATCAGCCTTTCGCTTTTGGCACTGAT
CTTGTCCTAGCAGAGTACTCTGACTCACAGAATGATGCAAATCAGCAAGCAATCAGCCCT
TTCTCTCCAGAAAGCAGAGAGCTTTTGGCTAGACTTGACGACCCTTTTTACTATGATATA
CTTGACCAGCTGATTTCTCTGAGCCATTGTTCTGCTCTCCCTCAGCCGTTCTTCGAGCCT
TCGCCGTGTGCCGAGAAGATGCAGACATGTTTCAAAGGATGAAGAAGCTGATGTTTTCTTA
GACGATTTCCAGCTGACATGTTTGATCAGGTTGATCCAATCCCAAGTCCTTAG

>G2720 Amino Acid Sequence (domain in AA coordinates: 10-114)

MEAKKEEIKKGPWKAEEDEVLINHVKRYGPRDWSSIRSKGLLQRTGKSCRLRWVNKLRPN
LKNGCKFSADEERTVIELQSEFGNKWARIATYLPGRTDNDVKNFWSSRQKRLARILHNSS
DASSSSFNPKSSSSSHRLKGKNVKPIRQSSQGFGLVEEEVTVSSSCSQMVPISSDQVGDEV
LRLPDLGVKLEHQPFAGFTDLVLAEYSDSQNDANQQAISPFSPESRELLARLDDPFYDI
LGPADSSSEPLFALPQPFEPSPVPRRCRHVSKDEEADVFLDDFPADMFDQVDPIPS*

>G2787 (142..1584)

TCTCAGAGCAAAAAACAAAAAAGAAAAAACCCTAAATCTAAATCTCACCTTCCA
CCTCTGTCTTTTTTTTTTTTGTCTTTTTTTTTTTTTTACTGTATCTTCTCTCTCTTGT
CTCTGCAAAAATCTCACATCCATGGATCCATCTCTTGGTGATCCTCATCATCTCTCAG
TTCACCCCTTTTCTCATTTTCCACCTCCAATCATCATCTTTAGGACCAATCCGTAC
AATAACCATGTCTCTTCCAACCGCAGCCGCAACGCAACGCAATCCCGCAACCGCAG
ATGTTTCAGTTATCTCCACATGTTTCAATGCCCCACCCTCCTTACTCCGAAATGATTGC
GCTGCGATTGCGGCGTTAAACGAACCGGATGGTTCGAGCAAGATGGCAATTTTCGAGATAC

ATCGAGAGATGTTACACCGGTTTAACTTCTGCTCATGCTGCTTTGTTGACTCACCATCTC
AAGACTTTGAAGACCAGTGGTGTCTTTCTATGGTTAAGAAATCTTACAAAATTGCTGGT
TCTTCTACTCCTCCTGCTAGTGTAGCTGTTGCTGCTGCTGCCGCCGCTCAAGGTCTCGAT
GTTCCCAGATCTGAGATTCTCCATTCAAGTAACAACGATCCCATGGCTTCTGGCTCTGCT
TCTCAGCCTCTGAAACGAGGTCGTGGTCTCTCTAAGCCTAAACCTGAATCTCAACCA
CAACCACTACAGCAACTTCCACCGACCAATCAAGTCCAGGCTAACGGACAGCCAATCTGG
GAACAGCAGCAAGTTCAATCACCTGTTCCGGTTCCGACTCCGGTTACAGAGTCGGCGAAG
AGAGGACCTGGTCTGTCACGAAGAAGCGTTCTGCTGCTCCTGCTACTGCACCAATCGTT
CAAGCTTCCGTTATGGCTGGAATTATGAAACGTAGAGGTAGACCACCGGGTCGTGAGCT
GCTGGGAGACAGAGGAAGCCCAAATCCGTTTCTTCTACTGCCTCTGTGTATCCTTATGTT
GCTAATGGTGCTAGACGCAGAGGAAGGCCCTAGGAGAGTTGTTGACCCTAGCAGTATTGTT
AGTGTGCTCCAGTAGGTGGTGAAAATGTGGCAGCGGTTGCGCCAGGGATGAAGCGTGGA
CGTGACGACCACCTAAGATTGGTGGTGTATCAGTAGGCTTATTATGAAGCCTAAGAGA
GGACGAGGACGTCTGTAGGTAGACCCAGAAAGATTGGAACATCAGTCACGACTGGGACA
CAAGATTCTGGAGAACTCAAGAAGAAGTTGATATTTTTCAAGAGAAAGTGAAGAAATTT
GTGAAGGTGTTGAAGGATGGAGTTACAAGTGAGAATCAAGCAGTGGTGCAAGCCATAAAA
GATCTGGAAGCACTAACAGTGACGGAGACCGTTGAGCCACAAGTTATGGAAGAAGTGACG
CCAGAGGAGACTGCAGCACCACAGACTGAAGCTCAACAACTGAAGCTGCTGAGACACAA
GGAGGACAAGAAGAAGGACAAGAAAGAGAAGGAGAAACACAGACCCAGACAGAAGCAGAG
GCAATGCAAGAAGCTCTGTTCTGAAGAATAATAATGATCTAGAAAACAACCTAGACATAA
TAGCCTTGGTGTGTTGGCGTTAGGAGTGTGTTTTTTTAGTTGTTTGGTGTGGAATCGC
ATCTTAAATTATATAAAAATCTATAAGGAATTTAATTTTTCTAGGTTTGTGCTGCA
GAAGAAGAAATAGTAGACTCGTTAATGGTGTGTTGTCGGTGTGCTTTAACCAAACCAT
AAGACGTGGCTGTAAATTAGCGATGTTTCTAGTCTTCCATCTTAAATAATCTCTTATTGC
GTCTGTGCCTTTGTTTTT

>G2787 Amino Acid Sequence (domain in AA coordinates: 172-192, 226-247, 256-276, 290-311, 245-366)

MDPSLGDPHPPQFTPFPHFPTSNNHPLGPNPYNHVVFPQPOTQTQIPQPQMFQLSPH
VSMPPHPYSEMICAIAALNEPDGSSKMAISRYIERCYTGLTSAHAALLTHHLKTLKTS
VLSMVKKSYKIAGSSSTPPASVAVAAAAAQGLDVPRSEILHSSNNDPMASGSASQPLKRG
RGRPPKPKPESQPQLQPLPTNQVQANGQPIWEQQQVQSPVPVPTPVTESAKRGPGRPR
KNGSAAPATAPIVQASVMAGIMKRRGRPPGRRRAGRQRPKSVSSTASVYPYVANGARRR
GRPRRVDPSSIVSVAPVGGENVAAPVAPGMKRRGRPPKIGGVISRLIMKPKRGRGPVG
RPRKIGTSVTTGTQDSGELKKKFDIFQEKVKEIVKVLKDGVTSENQAVVQAIKDLEALTV
TETVEPQVMEEVQPEETAAPQTEAQQTAAETQGGQEEGQEREGETQTQTEABAMQEALF
*

>G2789 (82..879)

CTTTAGGGACACCAAATCTATTCAACCTAAAAGCCTTCTTTTCCCCTATATTGACCAACT
TTTTAGCGAATCAGAAGAGGAATGGATGAGGTATCTCGTTCTCATACACCGCAATTTCTA
TCAAGTGATCATCAGCACTATCACCATCAAAACGCTGGACGACAAAACGCGGCAGAGAA
GAAGAAGGAGTTGAACCCAACAATATAGGGGAAGACCTAGCCACCTTTCTTCCGGAGAA
GAGAATATCAAGAAGAGAAGGCCACGTGGCAGACCTGCTGGTTCCAAGAACAAACCCAAA
GCACCAATCATAGTCACTCGCGACTCCGCGAACGCCTTCAGATGTACGTCATGGAGATA
ACCAACGCCTGCGATGTAATGGAAGCCTAGCCGTCTTCGCTAGACGCCGTCAGCGTGGC
GTTTGCGTCTTGACCGGAAACGGGGCCGTTACAAACGTCACCGTTAGACAACTGGCGGA
GGCGTCTGTCAGTTTACACGGACGTTTGAGATTCTTCTCTCTCGGGTTCGTTTCTTCT
CCACCGGCACCAACAGCTGCGTCTGGTTTAAAGGTTTACTTAGCCGGTGGTCAAGGTCAA
GTGATCGGAGGCAGTGTGGTGGGACCGCTTACGGCATCAAGTCCGGTGGTCTGTTATGGCA
GCTTCATTGGAACGCATCTTACGAGAGGCTGCCACTAGAGGAGGAGGAGGAACTGAA
AGAGAAATAGATGGAACGCGGCTAGGGCGATTGGAACGCAACGCAGAAACAGTTAATG
CAAGATGCGACATCGTTTATTGGGTCGCGCTCGAATTTAATTAACCTCTGTTTCGTTGCCA
GGTGAAGCTTATTGGGGAACGCAACGACCGCTTCTTAAGATAATATCATGTGATAATATA
AGTTTCGTCTTCTTATTCTTTTCACTTTTTTACCTTTTTTCACTTTCTTAGGTTTTGTTTT
AACGTTTGATTAATACCTGAAGGTTTTTGGAAAATTTTCGATCGGATAAAAGGATTTATG
TTGCGAGCCGAAACGCGGCC

>G2789 Amino Acid Sequence (domain in AA coordinates: 53-73, 121-165)
MDEVSRSHTPQFLSSDHQHYHHQNAQRQKRGREEGVEPNNIGEDLATFPSGEENIKRR

PRGRPAGSKNPKAPIIVTRDSANAFRCHVMEITNACDVMESLAVFARRRQRGVCVLTGN
GAVTNVTVRQPGGGVSLHGRFEILSLSGSFLPPAPPAAAGSLKVYLAGGQGVIGGSVV
GPLTASSPVVMAASFGNASYERLPLEEEEEETEREIDGNAARAIGTQTQKQLMQDATSF
GSPSNLINSVSLPGEAYWGTQRPSF*

>G31 (13..615)

CTTTTATAAGCAATGGCTCCAAGACAGGCGAACGGTAGAAGCATTGCCGTGAGTGAAGGC
GGCGGAGGGAAGACGATGACGATGACGACGATGCGGAAGGAAGTGCACCTTAGAGGTGTG
AGGAAGCGTCCATGGGGTAGATACGCGCGGAGATCCGTGACCCGGGAAAGAAAACCCGG
GTTTGGCTCGGGACATTTCGACACGGCGGAGGAAGCTGCAAGAGCTTACGACACCGCCGCT
AGAGAGTTTCGTGGCTCCAAAGCAAAGACTAATTTCCCTCTTCCCGGAGAGTCTACTACG
GTTAACGACGGTGGCGAGAACGATTCTTACGTCAACCGTACGACGGTGACGACGGCGCGT
GAGATGACGCGTCAGAGATTTCCGTTCGATGTCACCGGAGCGTAAAGTCGTCGGTGGT
TATGCTTCTGCTGGTTTTTTCTTCGATCCGTCAAGAGCTGCTTCGTTAAGAGCAGAGCTT
TCTCGGTTTGTCCGGTTCGGTTTGATCCGGTTAATATCGAGTTGAGTATTGGTATTCGA
GAAACCGTAAAGTTGAACCGAGAAGAGAATAAACCTGGATCTTAACCTAGCTCCACCG
GTGGTGGACGTTTAGATTTTTTTCTTCTTTTCATAATTTGTATTTTACATTGCCGAAAA
TAATTAATGTTTTCTTTAG

>G31 Amino Acid Sequence (domain in AA coordinates: TBD)

MAPRQANGRSIAVSEGGGKMTMTTMRKEVHFRGVRKRPWGRYAAEIRDPGKKTRVWL
TFDTAEBAARAYDTAAREFRGSKAKTNFPLPGESTTVNDGGENDSYVNRRTVTTAREMTR
QRFPPFACHRERKVVGGYASAGFFDPSRAASLRAELSRVCPVRFPVNIELSIGIRETVK
VEPRRELNLNLNAPPVVDV*

>G33 (20..757)

ATTCTCCCCAACCAAAATATGACCACAGAAAAAGAGAATGTCACTACGGCCGTGGCCGT
GAAAGACGGCGGAGAAAAGAGTAAGGAAGTGAGTGACAAGGGCGTAAAGAAGAGAAAGAA
TGTAACTAAGGCCCTGGCCGTGAATGACGGCGGAGAAAAGAGTAAGGAAGTGCGTTACAG
GGGTGTAAGGAGGAGACCATGGGGGAGATATGCTGCGGAGATCCGTGATCCGGTAAAGAA
AAAACGGGTCTGGCTCGGGTCCCTTCAACACGGGGGAGGAAGCCGCCAGAGCCTACGACTC
CGCTGCCATAAGGTTTCGAGGATCGAAAGCTACTACTAACTTCCCTCTAATCGGATACTA
TGGGATTCTTTCGGCGACGCCGTGAACAACAACCTTTCCGAGACGGTGAGTGATGGAAA
TGCCAACTCCCTCTCGTTGGAGACGATGGGAATGCTTTGGCTTCTCCGGTGAACAACAC
CCTTTCCGAAACGGCGCGTGATGGAACACTTCCATCGGATTGTACGACATGTTATCTCC
GGGGTGGCTGAAGCGGTTGCTGGATTTTTCTTAGATCTGCCTGAAGTTATTGCGTTGAA
AGAGGAGCTTGATCGAGTTTGTCTGACAGTTTGAGTCCATTGATATGGGGTTGACTAT
TGGTCTCAAAACCGCGGTGGAAGAGCTGAGACTTCTCCGGCGTGGATTGTAAGCTGCG
AATGGAACCGGATCTTGACCTCAACGCAAGTCCCTAAAGATTGATCTGATGTTGTTGTCC
TTGAATAAGTTTGTATCTTGTGCTCTTCTGATTGTCTGTACTTCTATTGGTTGATTGCG
TGCTTTTGGAGGACAAAACAAACATTTTTTTATGTATTAAGGTAATTGAACTATT
ATCGAAAAAAAAAAAAAAAAAAAAAAAAAAAAA

>G33 Amino Acid Sequence (domain in AA coordinates: 50-117)

MTTEKENVTTAVAVKDGGEKSKEVSDKGVKKRKNVTKALAVNDGGEKSKEVRYRGVRRRP
WGRYAAEIRDPVKKRVWLGSFNTGEEAARAYDSAAIRFRGSKATTNPLIGYYGISSAT
PVNNNLSETVSDGNANLPLVGDDGNALASPVNNLSETARDGTLPSDCHDMLSPGVAEAV
AGFFLDLPEVIALKEELDRVCPDQFESIDMGLTIGPQTAVEEPETSSAVDCKLRMEPDLD
LNASP*

>G342 (1..723)

ATGGACGTCTACGGCATGTCTTACCAGGACTTGCTTCGTATCGACGACCTTCTCGATTTC
TCCAACGACGAAATSTTCTCTTCCCTCTTCCACCGTCACTTCTCCCGCGCTTCTCCGCC
GCTTCTTCCGAAAACCTTTTCAGCTTTCTTCTTCCACCTACACTTCTCCTACTCTCCTC
ACCGACTTCACTACGATCTCTGCGTTCAGTGACGACGAGCTCATCTCGAATGGTTA
TCGCGATTGCTTGACGATTCACTTCTCGATTTCAGCAAAATCCTTTAACCATGACCGTT
AGACCCGAGATTTCACTTACCAGGAAAACCTAGAAGTCGCCGATCAAGAGCACCAGCACCT
TCCGTAGCTGGAACCTGGGCTCCGATGTCTGAATCAGAGCTTGTCACTCCGTCGCTAAA
CCTAAACCGAAGAAAGTCTACAACGCTGAATCGGTTACGGCGGATGGAGCGAGGCGGTGC
ACGCACTGTGCTCGGAGAAAACGCCACAGTGGAGAACTGGACCGCTTGGACCTAAACA
CTTTGTAACGCTTGTGGAGTTCGTTACAAATCAGGGAGGCTTGTACCGGAATACAGACCG
GCGTCGAGTCCGACGTTTGTATTGACTCAGCATTCGAACCTCTCATCGGAAAGTTATGGAG

CAAGCTCTCGGTGGTCACATGAGGCGGCACAGGACAGCCGTAACCACGATTAGCCCCGTT
GCAGCCACCCGAGAAGTAAGCAGAAACAGTACAGAGGAAGAGATTGAGATCAATATAGGC
CGTTCGATGGAACAGCAGAGGAAATATCTACCGTTGGATCTTAATCTACCAGCACCAGAA
GATGATCTAAGAGAGTCAAAGTTTCAAGGGATAGTATTCTCAGCAACACCAGCGTTAATA
GATTGTCATTACTAG

>G358 Amino Acid Sequence (domain in AA coordinates: 124-135, 188-210)

MGQDEVGSDQTQIIKGKRTKRQRSSSTFVVTAATTVTSTSSSAGGSGGERAVSDEYNSAV
SSPVTTDCTQEBEDMAICLIMLARGTVLPSPDLKNSRKIHQKISSENSSFYVYECKTCNR
TFSSFQALGGHRASHKKPRTSTEEKTRLPLTQPKSSASEEQNSHFVSGSALASQASNI
INKANKVHECSICGSEFTSGQALGGHMRRHRTAVTTISPVAATAEVSERNSTEEIEINIG
RSMEQQRKYLPLDLNLPAPEDDLRESKFQGI VFSATPALIDCHY*

>G360 (1..543)

ATGTGGAACCTTAACAAAATTGAAGAATTGGAGGATGATGATGAATCTTGGGAAGTCAAA
GCCTTTGAGCAAGACACTAAAGGCAACATCTCTGGTACCCTTGGCCTCCAAGATCTTAC
ACTTGCAATTTCTGCCGCCGTGAGTTCGGTCTGCTCAAGCCTTAGGCGGTCACATGAAT
GTCCACCGCCGTGACCGCGCCTCATCTAGGGCTCATCAAGGTTCCACCGTTGCGGCTGCG
GCTAGAAGCGGCCACGGGGGGATGTTACTCAATTCTGTGCTCCGCCGTTGCCTACAACG
ACACTTATAATACAACTCCACGGCGAGTAACATTGAAGGTTTGTCCATTTCTACCAACTG
CAAAACCCTAGTGGCATTTTTGGTAATTCTGGTGACATGGTGAATCTTTATGTAGAAGTT
CCTCTCGGCTTATTGAATATTTCGACAGGAGATGATGAGAGCATTGGCTCGATGAAAGAA
GCGACAGGAACATCAGTGATGAGCTTGATCTTGAAGTTTCGGCTAGGGCACCATCCACCG
TGA

>G360 Amino Acid Sequence (domain in aa coordinates: 42-62)

MWNPNKIEELEDDESWEVKAFEQDTKGNISGTTWPPRSYTCNFCRREFRSAQALGGHMN
VHRRDRASSRAHQGSTVAAAARSGHGMLLNSCAPPLPTTLIIQSTASNIEGLSHFYQL
QNPSGIFGNSGDMVNLYVEVPPRLIEYSTGDDESIGSMKEATGTSVDELDLELRLGHHP
*

>G362 (195..830)

ATAAAAAACCTTCATACAATATAAAATTTCTTTAGACATACAATATATTATACTATTAC
AGATGCAATGCATCATTAGTTACAACTATTAACTAAATATCCCCGCTCTCTCTTTGC
TATATAAGAAGATCATTTACACATCTCCTTAAGCAAATTAAACCCATCGATAAACACAT
ACGTTACACATATATGTCTATAAATCCGACAATGTCTCGTACTGGCGAAAGTTCTTCAG
GTTCTGCTCTCCGACAAGACGATAAAGCTATTCCGGCTTCGAAGTCTATCAGCGGCAGTCGTA
CGCCGGAATACACGACGGCGGAAAGCGTGAGCTCGTCCACAAACACGACGTCGTTAACAG
TGATGAAAAGACACGAGTGCCAATACTGCGGTAAAGAGTTTGCAAATTTCTCAAGCCTTAG
GAGGTACCAAAACGCTCACAGAAGGAGAGGTTGAAGAAGAAGAGGCTTCAGCTTCAAG
CTCGGCGAGCCAGCATCGGCTATTATCTACCAACCACCAACAACCCATAACGACGTCAT
TTCAGAGACAATACAAACGCCGTCGTATTGTGCATTTCTCCTCATGCACGTGAATAATG
ATCAGATGGGTGTGTACAACGAAGATTGGTCGTCGAGTCGTCGAGATTAAGTTCGGTA
ATAATGACACGTGCCAAGATCTTAATGAACAAAGCGGTGAGATGGGTAAAGCTGTACGGTG
TTCGACCGAACATGATTCAAGTTCAGAGAGATCTGAGTTCTCGTTCTGATCAGATGAGAA
GTATTAAGTTCGCTGGATCTTCATCTAGGTTTTGCCGAGATGCGGCATAACAAATTAAAG
AGAGATATATGATTAAGATTATGTACTATAGTGGCGTATTTTCATTGGGATCATGAAGG
GGAAAAACGAGACATATAGTATTCTTGATGCAATTTGAGTTTTGTAATTTATTAGGTT
TATGTATGTTTTTCGAAG

>G362 Amino Acid Sequence (domain in AA coordinates: 62-82)

MSINPTMSRTGESSSGSSDKTIKLFGEFELISGSRTPFITTAESVSSSTNTSLTVMKRH
ECQYCGKEFANSQALGGHQNAHKKERLKKRLQLQARRASIGYYLTNHQQPITTSFQRQY
KTPSYCAFSSMHVNNDQMGVYNEDWSSRSSQINFGNNDTCQDLNEQSGEMGKLYGVRPNM
IQFQRLSSRSQMSRINSIDLHLGFAGDAA*

>G364 (64..516)

AAGCTTGATATCGCCTCTCTCTAATCTCTCTTTCTCTCTCTATCTCTAAGAATATATAAA
GGTATGGACTACCAGCCAAACACATCCCTACGCTAAGCCTACCAAGTTACAAGAACCAC
CAACTAAACCTAGAAGTTGTTCTCGAGCCTTCTCCATGTCTTCTTCTCATCTTCTTCC
ACGAACTCATCATCATGTTTGGAGCAGCCTAGGGTATTCTCATGTAAGTATTGTCAAAGA
AAGTTTTACAGCTCTCAAGCTCTTGGTGGTCATCAAAACGCTCATAGCTTGAGAGAACC
TTAGCCAAGAAGAGTCGAGAACTCTTTAGATCCTCAAAACACTGTTGATTCTGATCAGCCT

TACCCGTTCTCCGGTCGCTTTGAGCTTTACGGCCGTGGCTACCAAGGATTTCTCGAAAGT
GGCGGCTCGAGGGACTTCTCCGCCCCCGGTGTGCCGGAGAGTGGTCTTGATCAGGATCAG
GAGAAGAGTCACCTTGACTTATCCTTAAGGCTCTAAAAGAATCTTATATTTTGTTAGTCT
ATATATTATCATATCAATTGTTAATCTTAAAATTGATTGTTTTACTTATTAGTCATTTC
TATTATCTGAAAGTTTTCTTTGTAAGTTGTAACATATGGTCTTAAATCAAATCCAAATTT
GATTTTGAAGATGGTACCTAATGCAGTAGTTAAATAAGTTAAAAAATGAAGGATCTAT
AATTCTCT

>G364 Amino Acid Sequence (domain in AA coordinates: 54-76)
MDYQPNLSRLSLPSYKNHQLNLELVLEPSSMSSSSSSSTNSSSCLEQPRVFSQNYCQRK
FYSSQALGGHQNAHKLERTLAKKSRELFRSSNTVDSQPYPFSGRFELYGRGYQGFLSEG
GSRDFSARRVPESGLDQDQEKSHLDLSRL*

>G365 (69..755)
CAATTCTTTTACTTTTCTTCTTTATATATTCTCTCTACGCTATAATATATATTACACA
GAATATACATGGAACCGTCCATCAAAGGAGATCAAGAAATGTTAAAAATCAAGAAACAAG
GTCATCAAGATCTTGAGTTGGGGTTGACCCCTTTTGTACGTGGAACCGCGACCTCATCAG
AGCTCAATCTCATCGATTCTTTCAAACCAGCTCATCATCGACTTCTCATCATCAGCACC
AGCAAGAACAATTGGCAGATCCGAGAGTGTCTCGTGAATTATTGTCAAAGAAAGTTCT
ATAGTTCAAGCGCTAGGCGGTACCAAAACGCTCATAAACGTGAGCGCACCTTAGCCA
AACGTGGACAGATTACAAGATGACTCTCTCCTTGCCTTCTTACGCGTTTGCCTTTG
GCCACGGTTCAGTCAGCAGATTTCGAAGCATGGCATCGTTACCATTACATGGCTCGGTGA
ATAACAGGTCAACGTTAGGGATTCAAGCTCATTCAACGATCCATAAGCCAGCTTCTTAG
GAAGACAAACGACGAGTTTAAGTCATGTTTTCAACAGAGCATTACCAGAAACCGACCA
TAGGAAAGATGTTGCCGAGAAATTCACCTTGAAGTCGCCGGAATAATAACAGTAACA
TGGTTGCTGCTAAGTTGGAGAGAATTGGACATTTCAAGAGCAACCAAGAAGATCATAATC
AGTTTAAGAAAATTGACTTGACTCTTAAGCTATGAGCTCTGCCATCTTCTTTTATGCTT
CATTATAACTTTTTTATCTCATCTTTGTTTGATATAATGATTGACGGCAGGGTGTGTT
AGAGTTTCACTAATGATCAAGTTGTACTTTTTATATATTTTATTGATACCTTGTGATGT
AATTCAATATTTTAGGTCGTGTTTTT

>G365 Amino Acid Sequence (domain in aa coordinates: 70-90)
MEPSIKGDQEMLKIKKQGHQDLELGLTLLSRGTATSSSELNLIDSFKTSSSSSTSHHQHQQE
QLADPRVFSQNYCQRFYSSQALGGHQNAHKRERTLAKRGQYYKMTLSSLPSSAFAGHG
SVSRFASMASLPLHGVSNNRSTLGIQAHSTIHKPSFLGRQTSLSHVFKQSIHQKPTIGK
MLPEKFHLEVAGNNNSNMVAAKLERIGHFKSNQEDHNQFKKIDLTLLK*

>G367 (1..708)
ATGGACGCTTCAATAGTTTCTCATCCACTGCTTTTCCATATCAAGATTCTCTAAACCAG
AGCATCGAAGACGAAGAAAGAGCGTTTATAATTCTAGTCACGAACCTAATCTCATCGAC
TGCATAGACGACACAACGAGTATCGTTAACGAATCTACAACATCCACAGAACAAGCTT
TTCTCATGCAACTATTGTCAAAGAACTTTCTATAGCTCACAAGCACTTGGTGGTCACCAA
AACGCACACAAGAGAGAGAGAACGTTGGCGAAGAGAGGACAACGTATGGCAGCGTCAGCC
TCAGCTTTTGGACATCCTTACGGTTTCTCTCCACTTCTTTCCACGGACAGTACAACAAC
CATAGGTCTTTAGGGATCCAAGCGCATTTCGATAAGCCACAAGCTAAGTTCTTATAACGGG
TTTGGTGGTCACTATGGTCAGATCAACTGGTCAAGACTTCCATTTGATCAACAACCAGCC
ATAGGTAAATTTCCCTCAATGGATAATTTTCATCATCATCATCATCAGATGATGATGATG
GCTCCTTCAGTAAATTCACGGTCCAATAACATCGATAGCCCAAGCAACACAGGACGGGTT
CTAGAAGGGTCACCGACTCTTGAACAATGGCACGGAGACAAAGGATTGTTGTTAAGTACA
AGTCATCATGAAGAGCAGCAGAACTTGACTTGTCCTCAAGCTTTGA

>G367 Amino Acid Sequence (domain in AA coordinates: 63-84)
MDASIVSSSTAFPYQDSLNLQSI EDEERDVHNSHLENLIDCIDDTSIVNESTTSTEQKL
FSCNYCQRTFYSSQALGGHQNAHKRERTLAKRGQMAASASAFGHPYGFSPLPFHGQYNN
HRSLGIQAHSISHKLSSYNGFGGHYQINWSRLPFDQQPAIGKFPMDNFHHHHQMMMM
APSVNRSNNIDSPSNTGRVLEGSPTLEQWHGDKGLLLSTSHHEEQKLDLSLKL*

>G373 (1..1854)
ATGGCGATTGAAACTCAGCTTCCCTTGCACGGTGACGGTGTGTGATGCGGTGTCAGGTG
AATCCTCCGTGAGAAGAGACTCTCACTTGTGGCACGTGCGTCACTCCATGGCACGTGCCG
TGTCTCTCCCGAATCACTCGCTTCTTCCACTGGAGAGTGGGAGTGTCCCGATTGCTCC
GGCGTTGTCTGTTCCCTCCGCCGTCCGGGTACCGGAAACGCTCGACCTGAATCTTCCGGT
TCAGTTCTCGTTGCTGCGATCCGTGCGATTACGGCTGATGAGACTTTAACCGAAGCTGAG

AAAGCCAAAAAAGGCAGAACTGATGAGTGGGGGTGGTGACGATGGTGTGATGAAGAA
GAGAAGAAGAAGTTAGAAATCTTTTGTCTATTGTCATTCAATTGCCAGAAAGACCTATC
ACGACACCGTGTGGGCACAAATTTCTGTTTGAATGTTTCGAGAAATGGGCAGTAGGTCAA
GGGAAGCTAACTTGTATGATATGCCGAAGCAAAATTCGAGACATGTGGCAAAAAATCCT
CGCATCAACTTAGCTCTAGTTTCTGCTATTCTGTTTAGCAAATGTTACCAAATGTTCTGTT
GAGGCAACTGCAGCCAAGGTTTCATCATATTATCCGCAACCAAGACCGTCTGAGAAAGCA
TTTACTACCGAGCGGGCAGTAAAACTGGGAAAGCTAATGCTGCTAGCGGTAAGTTTTTT
GTGACAATACCTCGTGATCATTTTGGTCCCATACCAGCTGAGAAATGATGTCACTAGAAAG
CAAGGTGTTTTGGTTGGAGAATCTTGGGAGGACAGGCAAGAGTGTAGGCAGTGGGGAGCT
CATTTCCCGCATATTGCTGGCATTGCCGGGCAATCAGCGGTGGAGCTCAGTCTGTGGCC
CTCTCTGGAGGTTATACGATGATGAGGATCATGGTGAATGGTTTCTCTACACAGGAAGT
GGTGGAAAGGATCTCAGTGGAAACAAAGAATTAAACAAGAAACAGTCGTCTGACCAGGCG
TTTAAAAACATGAATGAATCTCTAAGACTTAGTTGCAAAATGGGCTATCCTGTCCGAGTT
GTCAGGTCTTGAAGGAGAAGCGTTCTGCATATGCCCTGCTGAAGGTGTGAGATATGAT
GGGGTCTATCGAATTGAGAAGTGCTGGAGTAATGTTGGAGTACAGGGTTCTTTAAGGTC
TGTCGTTACCTGTTTGTAGATGTGACAATGAGCCAGCTCCATGGACCAGTGATGAGCAT
GGCGATCGTCCAAGACCGTTGCCAATGTTCCCGAGCTTGAGACTGCTGCTGACCTGTTT
GTGAGAAAGGAGAGTCCATCATGGGATTTTCGATGAAGCTGAGGGTCGTTGGAAATGGATG
AAGTCTCTCTGTTAGCAGAATGGCTTTGGATCCTGAGGAGAGGAAGAAGAATAAGAGA
GCAAAAAATACATATGAAGGCCAGACTTCTGAAAGAATTTAGTTGCCAAATCTGTCTGGGAA
GTGCTGAGTCTTCCAGTGACGACGCCTTGTGCACACAACCTCTGCAAAGCATGCTTAGAA
GCGAAGTTTGCTGGGATAACTCAACTGAGAGAGAGAAGCAATGGCGGACGTAAACTACGT
GCAAGAAGAACATCATGACCTGCCCTTGCTGCACGACGGATCTCTCCGAGTTTCTCCAA
AACC CGCAGGTGAACAGAGAGATGATGGAGATAATAGAGAATTTTAAGAAGAGTGAGGAA
GAGGCTGATGCATCCATTTCTGAAGAAGAAGAAGAATCCGAACCTCCAAC TAAGAAG
ATTAAGATGGATAACAACCTCTGTTGGTGGTAGTGGTACAAGTCTCTCAGCTTAA

>G373 Amino Acid Sequence (domain in AA coordinates: 129-168)
MAIETQLPCDGDGVCMRCQVNPSEETLTGCTCVTPWHVPCLLPESLASSTGEWECPCDS
GVVVP SAAPGTGNARPESG SVLVAAIRAIQADETLTEAEKAKKRQKLMSGGGDDGVDEE
EKKKLEIFCSICIQLPERPITTPCGHNFCLKCFEKWAVGQGLTCMICKRSKIPRHVAKNP
RINLALVSAIRLANVTKCSVEATAAKVHHIIRNQDRPEKAFTTERAVKTGKANAASGKFF
VTIPRDFHGP IPAENDVTRKQGVLVGESWEDRQECRQWGAHFPHIAGIAGQSAVGAQSV
LSGGYDDDEDHGEWFLYTGSGGRDLSGNKRINKKQSSDQAFKNMNE SLRLSCKMGYPVRV
VRSWKER SAYAPAEGVRYDGVYRIEKCWSNVGVQGSFKVCRYLFVRC DNEPAPWTSDEH
GDRPRPLPNVPELETAADLFVRKESPSWDFDEAEGRWKWMKSPVSRMALDPEERKKNKR
AKNTMKARLLKEFSCICREVLSPVTTPCA HNFCKACLEAKFAGITQLRERSNGGRKLR
AKKNIMTCPCCTTDLSEFLQNPQVNREMMEI IENFKKSEEEADASISEEEEESEPPTKK
IKMDNNSVGGSGTSLSA*

>G396 (1..957)
ATGGGGGAAAGAGATGATGGGTTGGGTTTGAGTCTAAGCTTGGGAAATAGTCAACAAAA
GAACCATCTCTGAGGTTGAATCTTATGCCGTTGACAACCTTCTTCTTCTTCTTCTGTTT
CAACACATGCACAATCAGAATAACAATAGCCATCCCCAGAAGATTATAACATCTCTTGG
ACTCATCTGTTTCAATCTTCTGGGATTAAACGTACAACCTGCAGAGAGAACTCCGACGCC
GGGTCAATTTCTAAGAGGTTTCAACGTGAACAGAGCTCAGTCTTCTGTTGGCGGTAGTGGAC
TTGGAAGAAGAAGCCGCCGCTCGTCTCGTCTCCAAACAGCGCCGTTTCGAGTCTGAGTGA
AATAAAAGGGATCTTGCGGTGGCGAGAGGAGGAGATGAAAACGAGGCGGAGAGAGCTTCT
TGCTCAGCGGAGGGGGAAGCGGTGGTAGCGACGATGAAGACGGCGGAAACGGCGACGGA
TCAAGGAAGAACTACGGTTATCGAAGGATCAAGCTCTTGTCTCGAGGAGACTTTTAAA
GAACATAGCACTCTTAATCCGAAGCAAAAGCTGGCTCTAGCAAAAACAGTTGAATCTAAGG
GCAAGACAAGTTGAAGTGTGGTTTTCAGAACCGTAGGGCAAGGACGAAGCTGAAACAAACG
GAGGTTGATTGTGAGTATTTAAAGAGATGTTGCGATAATCTGACCGAGGAGAATCGACGG
CTGCAGAAAGAAGTGTCTCGAGCTGAGGGCGTTGAAGTTGTCTCCACATCTCTACATGCAC
ATGACTCCTCTACTACTCTCACCATGTGCCCTTCTTGCGAACGTGTCTCTCTCTCTGCC
GCCACTGTGACCGCTGCTCCTTCCACTACTACTCTCTACGGTGGTGGGGCGGCCAAGT
CCACAGCGATTAACTCCTTGGACTGCTATTTCTCTCCAGCAAAAATCAGGTCGCTAG
>G396 Amino Acid Sequence (domain in AA coordinates: 159-220)
MGERDDGLGLSLSLGNSQQKEPSLRNLNLMPLTSSSSSSSFQHMHNQNNNSHPQKIHNISW

THLFQSSGIKRTTAERNSDAGSFLRGFNVNRAQSSVAVVDLEEEAAVVSSPNSAVSSLSG
NKRDLAVARGGDENEAEERASCSRGGSGGSDDEDGNGDGSRRKLRLSKDQALVLEETFK
EHSTLNPQKLALAKQLNLRRARQVEVWFQNRRTKLKQTEVDCEYLKRCCDNLTEENRR
LQKEVSELRLKLSPLHYMHMTPTTLTMCPSCERVSSAATVTAAPSTTTTPTTVGRPS
PQRLTPWTAISLQKSGR*

>G431 (1..1149)

ATGGAGAGTGGTTCCAACAGCACTTCTTGTCCAATGGCTTTTGCCGGGGATAATAGTGAT
GGTCCGATGTGTCTTATGATGATGATGATGCCGCCCATCATGACATCACATCAACATCAT
GGTCATGATCATCAACATCAACAACAAGAATCATGGTTATGCATATCAGTCACACCAC
CAACAAAGTAGTTCCCTTTTCTTCAATCACTAGCTCCTCCCCAAGGAACATAAGAACAAA
GTTGCTTCTTCTTCTTCT
CATAACGAAATCGTTGACAGGAGGAATCAACCCTTGCTCCTCTTCTCTCTCTCTCTCTCT
GTCAAGGCCAAGATCATGGCTCATCCTCACTACCACCGCCTCTTGCCGCTTATGTCAAT
TGTCAGAAGGTTGGAGCACCACCGAGGTTGTGGCGAGGCTGGAGGAGGCATGCTCGTCT
GCCGCAGCCGCAGCCGCATCTATGGGGCCAACAGGGTGTCTTGGTGAAGATCCAGGGCTT
GATCAATTCATGGAAGCTTACTGTGAAATGCTCGTTAAGTATGAGCAAGAGCTCTCCAAA
CCTTTCAAGGAAGCTATGGTCTTCTTCAACGTGTGAGTGTCAATTCAAATCCCTCTCT
CTATCTCACCTTCTCTCTTCTCTCGGTTATGGAGAGACAGCAATTGATAGGAACAATAAT
GGGTCATCCGAGGAAGAAGTCGATATGAACAATGAATTTGTAGATCCACAAGCTGAGGAT
AGAGAGCTTAAAGGACAGCTCTTGCGCAAGTACAGTGAGTTACTTAGGGAGCCTCAAGCAA
GAGTTCATGAAGAAGAGGAAGAAAGGAAAGCTCCCTAAAGAAGCTCGTCAACAACCTGCTT
GATTGGTGGAGCCGTCACTACAAATGGCCTTACCCTTCGGAGCAACAAAAGCTCGCCCTT
GCGGAATCAACGGGGCTGGACCAGAAACAGATAAACAATTGGTTCATAAACCAGAGGAAA
CGGCATTGGAAGCCGTGCGAGGACATGCAGTTTGTAGTAATGGACGCAACACATCCTCAC
CATTACTTCATGGATAATGTCTTGGACAATCCTTTCCCAATGGATCACATCTCCTCCACC
ATGCTTTGA

>G431 Amino Acid Sequence (domain in AA coordinates: 286-335)

MESGNSSTSCPMFAFAGDNSDGPMPMMMMPPIMTSHQHGHGHDHQHQQQEHDGYAYQSHH
QQSSSLFLQSLAPPQGTKNKVASSSSPSSCAPAYSLMEIHNEIVAGGINPCSSFSSSAS
VKAKIMAHPHYHRLLAAYVNCQKVGAPPEVVARLEEACSSAAAAAASMGPTGCLGEDPGL
DQFMEAYCEMLVKYEQLSKPFKEAMVFLQRVECQFKSLSLSSPSSFSGYGETAIDRNNN
GSSEEEVDMMNEFVDPQAEDELKQQLLRKYSGLSLKQEFMKRKRKGLPKPEARQQLL
DWWSRHYKWYPYSEQQKLALAEESTGLDQKQINNWFINQRKRHWKPSMEDMQFVVM DATHPH
HYFMDNVLDNPFPMDHISSTML*

>G479 (1..1128)

ATGGAGATGGGTTCCAACCTCGGGTCCGGGTTCATGGTCCGGGTGAGGCAGAGTCCGGTGGT
TCCTCCACTGAGTCATCCTCTTTTCAAGTGGAGGGCTCATGTTTGGCCAGAAGATCTACTTC
GAGGACGGTGGTGGTGGATCCGGGTCTTCTTCTCAGGTGGTCTGTTCAAACAGACGTGTC
CGTGGAGGCGGGTCCGGTTCAGTCCGGTTCAGATACCAAGGTGCCAAGTGAAGGTTGTGGG
ATGGATCTAACCAATGCAAAAGGTTATTACTCGAGACACCGAGTTTGTGGAGTGCACCTCT
AAAACACCTAAAGTCACTGTGGCTGGTATCGAACAGAGGTTTTGTCAACAGTGCAGCAGG
TTTCATCAGCTTCCGGAATTTGACCTAGAGAAAAGGAGTTGCCGAGGAGACTCGTGGT
CATAATGAGCGACGAAGGAAGCCACAGCCTGCGTCTCTCTCTCTCTCTCTCTCTCTCTCT
GGGAGGATCGCACCTTCTGCTTACGAAAATGGTGTGCTGGAATGAATGGAAGCTTTCTT
GGGAACCAAGAGATAGGATGGCCAAGTTCAAGAACATTGGATACAAGAGTGATGAGGCGG
CCAGTGTCTGTCACCGTCATGGCAGATCAATCCAATGAATGTATTTAGTCAAGGTTCAATT
GGTGGAGGAGGACAAAGCTTCTCATCTCCAGAGATTATGGACACTAACTAGAGAGCTAC
AAGGGAATTGGCGACTCAAACCTGTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT
GACA
TCAGGTTTTGGCCCGATGACGGTTACAATGGCTCAACCACCACCTGCACCTAGCCAGCAT
CAGTATCTGAACCCGCCCTTGGGTATTCAAGGACAATGATAATGATATGTCTCTCTCTCTCT
AATTTAGGTCGATACACCGAGCCAGATAAATTGTGATGAGTGTGACGACGCAATGGGT
GAGTTCGAGTTATCTGATCACCATCATCAAAGTAGGAGACAGTACATGGAAGATGAGAAC
ACAAGGCTTATGACTCTTCTTCTCACCATAACCAACTGGTCTCTCTGA

>G479 Amino Acid Sequence (conserved domain in AA coordinates:70-149)

MEMGNSGPGHGPQAESGSSSTESSFSGGLMFGQKIYFEDGGGSGSSSSGGRSNNRV
RGGGSGSQSQIPRCQVEGCGMDLTNAKGYYSRHRVCGVHSKTPKVTVAGIEQRFCQCSR

FHQLPEFDLEKRSRRRLAGHNERRRKQPASLSVLASRYGRIAPSLYENG DAGMNGSFL
GNQEIGWPSRTLDTRVMRRPVSSPSWQINPMNVFSQGSVGGGTSFSSPEIMDTKLESY
KGIGDSNCALLSNPHQPHDNNNNNNNNNNNTWRASSGFGPMTVTMAQPPPPAPSQH
QYLNPPWVFKDNDNDMSPVLNLGRYTEPDNCQISSGTAMGEFELSDHHHQSRRQYME DEN
TRAYDSSSHHTNWSL*

>G546 (1..588)

atgactcgaccgtcaagattacttgagacggcgccgcccaccaccacaaccgtcggaggag
atgatcgacggaatccgacatgggtggtgatcttgcggctcttcttgcgctcttatc
tgcgttgctggtctcgcgcgcgtcgtagatgcgcttggtccggcggtttacagccgga
ggagattcgccgtcaccgaacaaaggcttgaaaaagaaagctcttcagtctcttccaaga
tcactttccacgcccgcggaatcaacctccggcgccgcgctgaagagggagactcgacg
gaatgtgctatttgcctcactgacttcgcccgcggtgaagaaataagagtgccttctctt
tgtgtgcattcttccacgtggagtgtattgacaaatggctagtttctaggtcttctgt
ccttcttgcgcaggattcttacgcgggtgagatgtgaccggtgtggtcatgcttctacg
gcggagatgaaagatcaagctcatcgctcatcaacatcaccaacactcttctactaccatt
cctacgtttcttctcttaa

>G546 Amino Acid Sequence (domain in AA coordinates:114-155)

MTRPSRLLETAAPPPQPSEEMIAAESDMVVILSALLCALICVAGLAADVRCALRRFTAG
GDSPPNKGKLLKALQSLPRSTFTAAESTSGAAAEEDSTECALCLTDFADGEEIRVLPL
CGHSFHVBCIDKWLVSRSRSCPSRRILTPVRCDCRGHASTAEMKDQHRHQHHQHSSTTI
PTFLP*

>G551 (1..708)

ATGGAGTGGTCAACAACGAGCAACGTAGAAAACGTGAGAGTAGCTTTCATGCCACCGCCA
TGGCCGGAGTCTAGTTCCTTTAACTCGCTCCACAGCTTCAACTTTGATCCTTACGCAGGA
AATTCATATACGCCTGGCGATACACAAACCGGACCGGTATCTCTGTACCGGAATCAGAA
AAGATCATGAATGCGTACCGATTTCGAACAACAACATGAGATGATAAAAAAGAAGAGA
CTAACGAGTGGACAATTAGCTTCACTTGAGCGAAGTTTTCAAGAAGAGATCAAATTAGAT
TCAGACAGGAAGGTGAAGCTGTCGAGAGAGCTCGGTCTGCAGCCACGTCAGATAGCAGTT
TGTTTCCAAAACCGCCGTGCACGGTGGAAGGCGAAGCAGCTTGAGCAGTTGTACGACTCG
CTTAGACAAGAGTACGACGTCGTTTCTAGGGAGAAACAAATGTTACACGATGAGGTGAAG
AAGCTGAGAGCTTTACTAAGAGACCAGGGTTTGATCAAGAAGCAAATCTCTGCCGGGACC
ATCAAAGTTTCCGGTGAGGAAGACACGGTGAGATTTCATCGGTGGTGGTAGCTCATCCA
AGAACGGAGAATATGAACGCAAATCAAATCACCGGAGGGAATCAAGTTTACGGTCAATAC
AACATCCGATGCTGGTTGCTTCCTCTGGCTGGCCGTCATACCCCTGA

>G551 Amino Acid Sequence (conserved domain in AA coordinates:73-133)

MEWSTSNVENVRVAFMPWPPESSSFNSLHSFNFDYPAGNSYTPGDTQTGPVISVPESE
KIMNAYRFPNNNNEMIKKRLTSGQLASLERSFQEEIKLSDSRKVKLSRELGLQPRQIAV
WFQNRARRWKAKQLEQLYDSLQREYDVVSREKQMLHDEVKKLRALLRDQGLIKKQISAGT
IKVSGEEDTVEISSVVVAHPRTENMNANQITGGNQVYGQYNNPMLVASSGWPSYP*

>G578 (1..978)

ATGCATAGTTGAATGAAACAGTAATTCCTGATGTTGATTACATGCAGTCTGATAGAGGG
CATATGCATGCTGCTGCCTCTGATTCCAGTGATCGATCAAAGGATAAGTTGGATCAAAAG
ACCTTCGTAGGCTTGCTCAAAATCGTGAGGCAGCAAGAAAAGCAGATTGAGGAAGAAG
GCGTATGTTGAGCAGCTGGAAGATAGTCGATTAAAGCTGACTCAAGTTGAGCAGGAGCTG
CAAAGAGCAAGACAGCAGGGAGTTTTCATCTCAAGTTGAGGAGACCAAGCTCATTCTACT
GGTGGCAATGGTGGGGCTTTGGCATTGATGCAGAACACTCACGATGGCTTGAAGAAAAG
AACAGGCAAATGAACGAGCTGAGATCTGCCCTGAATGCTCATGCAGGTGATACTGAGCTC
CGGATAATTGTGGATGGAGTGATGGCTCACTATGAGGAGCTTTTCAGGATTAAGAGCAAT
GCATCTAAGAATGATGTCTTCCACTTGTATCTGGAATGTGGAAAACACCAGCTGAGCGA
TGTTTCTTGTGGCTTGGCGGGTTCCTGCATCCGAACCTTCTCAAGCTTCTTGGCAATCAG
CTAGAGCCCATGACAGAACGACAGGTAATGGGCATCAATAGCTTGCAGCAGACGTCGCAG
CAGGCAGAAGATGCTTTATCTCAAGGGATGGAGAGTTTACAGCAATCCCTAGCTGATACT
TTATCCAGTGGAACCTCTTGGTTCCAGTTCATCGGATAATGTTCGCGAGCTACATGGGTCAG
ATGGCCATGGCAATGGGCAAGTTAGGCACCTCGAAGGATTATACGCCAGGCTGATAAC
TTGAGGCTGCAAACTACAACAGATGCTTCGAGTATTAACAACACGTCAGTCAGCTCGT
GCTCTTCTTGTCTATACAGATTATTCATCTCGATTACGTGCTCTTAGTTCTTGTGGCTT
GCCCCGCCAAGAGAGTGA

>G578 Amino Acid Sequence (domain in AA coordinates 36-96)
MHSINETVLPDQVDMQSDRGHMHAAASDSSDRSKDLQKTLRRLAQNREAARKSRLRKK
AYVQQLSRLKLTQVEQELQRRARQQGVFISSSGDQAHSTGGNGGALAFDAEHSRWLEEK
NRQMNELRSALNAHAGDTELRIIVDGVMAHYEELFRIKSNASKNDVFHLLSGMWKTPAER
CFLWLGGFPSSSELLKLLANQLEPMTERRQVMGINSLQOTSQQAEDALSQGMESLQQLADT
LSSGTLGSSSSDNVASYMGQMAMAMGKLGTLGFIHQADNLRLQTLQQLRVLTTTRQSAR
ALLAIHDYSSRLRALSSLWLARPRE*

>G596 (168..1121)
TAATTTCTCTACTTCAGATTTTTTCTCCTTAGATTAAATTTAGTTATTGTACATC
CCTCAAGCTAAGATTCTGGTTTTGTGAGTTGAGTGGATGAGAAGAGGAGAGATTAACTAA
ATTAGGGTTTCAATTGTTTTACTTTTTGTTTGCTTTTATATCAAGTAATGGATCAGGTCT
CTCGCTCTCTTCTCCACCTTTTCTCTCAAGAGATCTCCATCTTCACCCACACCATCAAT
TCCAGCATCAGCAGCAGCAGCAACAGAATCACGGCCACGATATAGACCAGCACCGAA
TCGGTGGGCTAAAACGTGACCAGATGCTGATATCGATCCCAACGAGCACTCTTCAGCCG
GAAAAGATCAAAGTACTCCTGGCTCCGGTGGAGAAAGCGGCGGAGGAGGAGGAGATA
ATCACATCAGGAGAAGGCCACGTGGCAGACAGCGGGATCTAAGAACAAACCAAAACCGC
CAATCATCATCACTCGAGACAGCGCAAACGCTCTCAAACTCATGTTCATGGAAGTAGCAA
ACGGATGTGACGTGATGGAAGTGTACCGTCTTCGCTCGCCGTCGCCAACGTGGCATCT
GCGTTTTGAGCGGAAACGGCGCGTTACCAACGTTACCATAAGACAACAGCTTCAGTAC
CTGGTGGTGGCTCATCTGTCTTAACCTTACACGGACGTTTCGAGATTCTTTCTCTCTCGG
GATCATTCCTTCTCTCCGGCTCCACCAGCTGCGTCAGGTCTAACGATTACTTAGCCG
GTGGTCAGGACAGGTTGTTGGAGGAAGCGTGGTGGTCCACTCATGGCTTCAGGACCTG
TAGTGATTATGGCAGCTTCGTTTGGAAACGCTGCGTATGAGAGACTGCCGTTGGAGGAAG
ACGATCAAGAAGAGCAAACAGCTGGAGCGGTTGCTAATAATATCGATGGAAACGCAACAA
TGGGTGGTGGAAACGCAACGCAAACTCAGACGCAGCAGCAACAGCAACACAGTTGATGC
AAGATCCGACGTCGTTTATACAAGGGTTGCCTCCGAATCTTATGAATTCGTTCAATTGC
CAGCTGAAGCTTATTGGGGAACTCCGAGACCATCTTTCTAAATCGCGAAGAAAAACAAG
TTAGATACGTTTCGTTGTTTTTAATTTATAATCTCTCTCTGTCAAGTTTAAATTTCTTT
TTCTTCTCTTTGTTTTCTAAAGATAATTGTAGTCTTTGACGAAGATTCGTGGTACGTAT
GAATCGAAGAGAATCGTTTTGGTCATGGGATTGCTCGATCTATTAGGTTTGAGAGGGGT
TTGTGTTTTGCGTTGACTAGCAGATTATAAAATTGTTGATTTTCGAGTTTTTATTTTCAT
GTGTTGGTGATAAA

>G596 Amino Acid Sequence (domain in AA coordinates: 89-96)
MDQVSRSLPPFLSRDLHLHPHHQFQHQQQQQQNHGHDIDQHRIGGLKRDRDADIDPNE
HSSAGKDQSTPGSGGESGGGGDNHITRRPRGRPAGSKNPKPPIITRDSANALKSHV
MEVANGCDVMESVTVFARRRQRGICVLSGNGAVTNVTIRQPASVPGGSSSVNLHGRFEI
LSLSGSFLPPPAPPAASGLTIYLAGGQQQVVGSSVVGPLMASGPVVIMAASFGNAAYERL
PLEDDQEEQTAGAVANNIDGNATMGGGTQTQTQTQQQQQQQLMQDPTSF IQGLPPNLMN
SVQLPAEAYWGTTPRPSF*

>G617 (59..1141)
CAGATCTGTTCTTTACACCAAATTGAGTACTGAAGATCTTGTGAGTGAATTAAAGAGAT
GAGATCAGGAGAATGTGATGAAGAGGAGATTCAAGCAAAGCAAGAAAGAGATCAAAATCA
AAATCATCAAGTAAACTTAAACCACATGTTGCAACAACAACAGCCGAGTTCGGTATCATC
TTCAAGGCAATGGACTTCAGCTTTTAGGAATCCAAGAATCGTTTCGAGTCTCAAGAACATT
CGGTGGCAAAGACAGACACAGCAAAGTATGTACAGTCCGTGGTCTTCGAGACCGGAGGAT
AAGGTTGTCCGTACCTACAGCTATTCAACTCTACGACCTTCAAGATCGATTAGGGCTGAG
TCAGCCAAAGCAAGTCATTGATTTGGTTACTCGAAGCAGCAAAGATGACGTAGACAAGCT
ACCTCCTCTACAATCCCACATGGATTTAACCAGATGTATCCAATCTCATCTTCGGAAA
CTCCGGGTTTGGAGAATCTCCATCTTCAACTACATCAACAACGTTTCCAGGAACCAATCT
CGGGTTCTTGAAAATTGGGATCTTGGTGGTCTTCAAGAACAAGAGCAAGATTAAACCGA
TACAACACTACGACCCAAAGAGAAAGTTTGTATCTTGATAAAGGAAAATGGATCAAAAACGA
CGAAGATAGTAACTAAGATCATCAAGGGTTTAACACCAATCATCAACAACAATTTCTCT
GACCAATCCGTACAACAACACTTCAGCTTATTACAACCTTGGACATCTTCAACAATCGTT
AGACCAATCTGGTAATAACGTTACTGTGCAATATCTAATGTTGCTGCTAATAATAACAA
TAATCTCAATTTGCATCTCTCTCTCGTCTGCCGAGATGGATCTCAGCTTTTTTTCGG
TCCTACTCTCTCCGGCAATGAGCTCTCTATTCCCGACATACCCTTCGTTCTTGGAGCTTC
TCATCATCATCATGTCTCGATGGAGCCGGTCATCTTCAGCTCTTTAGCTCGAATTCAAA

TACCGCATCGCAGCAACACATGATGCCGGGTAATACGAGTTTGATTAGACCATTTCATCA
TTTGATGAGCTCGAATCATGATACGGATCATCATAGTAGCGATAATGAATCAGATTCTTG
AATGATTTTATATATCTACACTATACATTGAAAATGTTATATGTATACGTATTCTTCTAT
ATTTTGATATATATGCGTATTGTTGGATTGGTTTATGTATCT

>G617 Amino Acid Sequence (domain in AA coordinates: 64-118)
MRSGECDEEEIQAKQERDQNHQVNLNHLQQQQPSVSSSRQWTSAFRNPRIVRVSR
FGGKDRHSKVCTVRGLRDRRIRLSVPTAIQLYDLQDRLGLSQPSKVIDWLLEAAKDDVDK
LPPLQFPHGFMQYPNLI FGN SGFGESPSSTTSTTFPGTNLGFLENWDLGGSSRTRARLT
DTTTTQRESFDLDKGKWKINDENSNDHQGFNTNHQQQFPLTNPNNTSAYYNLGHLOQS
LDQSGNNVTVAISNVAANNNNNLNLHPPSSSAGDGSQLEFFGPTPPAMSSLFPTYPSFLGA
SHHHHVVDGAGHLQLFSSNSNTASQQHMPGNTSLIRPFHLMSSNHDTDHSSDNESDS
*

>G620 (40..666)
GAATTGAACCTGGACCAGCACAGCAACAACCCCAACCCCAATGACCAGCTCAGTCATAGTA
GCCGGCGCCGCGGTGACAAGAACAATGGTATCGTGGTCCAGCAGCAACCACCATGTGTGGCT
CGTGAGCAAGACCAATACATGCCAATCGCAAACGTCATAAGAATCATGCGTAAACCTTA
CCGTCTCAGCCAAAATCTCTGACGACGCCAAAGAAACGATTCAAGAATGTGTCTCCGAG
TACATCAGCTTCGTGACCGGTGAAGCCAACGAGCGTTGCCAACGTGAGCAACGTAAGACC
ATAACTGCTGAAGATATCCTTTGGGCTATGAGCAAGCTTGGGTTTCGATAACTACGTGGAC
CCCCTACCGTGTTTCATTAACCGGTACCGTGAGATAGAGACCGATCGTGGTTCTGCACCT
AGAGGTGAGCCACCGTCCGTTGAGACAAACCTATGGAGGAAATGGTATTGGGTTTCACGGC
CCATCTCATGGCCTACCTCCTCCGGGTCCTTATGGTTATGGTATGTTGGACCAATCCATG
GTTATGGGAGGTGGTCCGTACTACAAAACGGGTCGTCCGGTCAAGATGAATCCAGTGTT
GGTGGTGGCTCTTCGTCTTCCATTAACGGAATGCCGGCTTTTGACCATATATGGTCAGTAT
AAGTGAAGAAGGAGTTATTCTTCATTTTTATATCTATTCAAACATGTGTTTCGATAGAT
ATTTTATTTTATGTCTTATCAATAACATTTCTATATAATGTTGCTTCTTTAAGGAAAAG
TGTTGTATGTCAATACTTTATGAGAACTGATTTATATATGCAAAT

>G620 Amino Acid Sequence (domain in AA coordinates: 20-118)
MTSSVIVAGAGDKNNGIIVVQQPPCVAREQDQYMPIANVIRIMRKTLP SHAKISDDAKET
IQECVSEYISFVTGEANERCQREQRKTITAEDILWAMSKLGFNDYVDPLTVFINRYREIE
TDRGSALRGEPPSLRQTYGGNGIGFHGPSHGLPPPGPYGYMLDQSMVMGGGRYYQNGSS
GQDESSVGGSSSSSINGMPAFDHYGQYK*

>G625 (151..1137)
AATCGACCATTCAACGATGACATTCAAACACTCTTCAGTTTCCCTTCCTTCTTGATTCT
GTCCTCTCCACTATTTTCTCAATTTCTTTAATCTCTCTCTTCTCTCTACTTCTCTCT
TCCTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT
AACCATCTGGAAGATAATAACCAAAACCTAACCATAATAATCTCAATCCGATTCCACC
ACCGACTCATCAACTTCTTCCGTCAACGCAAACGCAAAGGCAAAGGTGGTCCGGACAAC
TCCAAGTTCCGTTACCGTGGCGTTCGACAAAGAAGCTGGGGCAAATGGGTCCGGGAGATC
CGAGAGCCACGTAAGCGCACTCGCAAGTGGCTTGGTACTTTCGCAACCGCGGAAGACGCC
GCACGTGCCTACGACCGGGCTGCCGTTTACCTATACGGGTACAGTGCTCAGCTCAACTTA
ACCCCTTCGTCTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT
CCTTCCACCTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT
GCCACCGTAGGAGGAGGACCAACTTTGGTCCGTACGGTATCCCTTTTAAACAACAACATC
TTCCTTAATGGTGGGACCTCTATGTTATGCCCTAGTTATGGTTTTTTTCCCTCAACAACAA
CAACAACAAAATCAGATGGTCCAGATGGGACAATTTCAAACACCAACAGTATCAGAATCTT
CATTCTAATACTAACAATAACAAGATTTCTGACATCGAGCTCACTGATGTTCCGGTAACT
AATTGCACTTCGTTTCATCATGAGGTGGCGTTAGGGCAGGAACAAGGAGGAAGTGGGTGT
AATAATAATAGTTTCGATGGAGGATTTGAACTCTCTAGCTGGTTCCGGTGGGTTCGAGTCTA
TCAATAACTCATCCACCGCGGTTGGTTGATCCGGTATGTTCTATGGGTCTGGATCCGGGT
TATATGGTTGGAGATGGATCTTCGACCATTGCGCTTTTGGAGGAGAAGAAGATATAGT
CATAATTGGGGGAGTATTTGGGATTTTATTGATCCCATCTTGGGGGAATTTCTATTAATTT
GTTTTTGTGGAAGATCATATTATATACGATGAGCATCCCTAAGGTCGGTCAAGAGCATTG
GAGATTCAATTGTTGAGAGGAATCAAAGAGATTGCATTCTATGAGGAGCTCTGCATGCAAA
ATTTTGGAGGATTTTTTACTACCTATAGAGATAAATAAGAGGGTATTTTTATTATTTTT
TTGAAGATTTTTATTTCAGGAATTCGTAAAAGAGATACGGTTCCAATAAAGTATGTA
TATGTGGAAGAGAATCGGAGGAGATGGTGGAAAGTTGTATGGGAATTTTATTGGTTCAAC

ACTCCTCCTCCTCCAACGAATTGTAGCTATAGGTTGTTTGGATTTGATCTCACAAGCAAT
TCTCCTGCTCCAATCCCTCAAGACAAGCAACCGATGGATACTTGTGGAGCTGCCAAGTGT
CAAGAACCCATCACTCCAACCTCAATGAGTGAGCAGAAGAAGCAACAAACATCAAGAAGT
CGAACTAAAGTGCAAATGCAAGGCATTGCGGTTGGTCGTGCGGTTGATTTAACTGTTG
AAATCTTACGATGAAGTATTGATGAGCTTGAGGAGATGTTTGAGATTCAAGGACAGCTT
CTTGCCCAGACAAATGGATCGTTGTCTTCACTGATGATGAAGGAGATATGATGCTTGCT
GGTATGATCCGTGGAATGAGTTTGTCAAGATGGCAAAGAAGATATTTATATATTCGAGC
GATGAGGTTAAGAAAATGACAACGAACTGAAGATTTCTTCGTGCTTAGAGAATGAGGAA
TATGGTAATGAATCATTGCAAAATCGTAGTAGGGGGTGAGAGTTTGTAGCTGTTAATTAAG
GTTAATTCGCGACGTCGTTTGTAGTGCCTAAGTGTCTAAAGACTTTTTTTTAGTCTGTG
TATATAAAGTCTTGTCTCTTTTTCATGTCAATTTTCAAGTTGGCGATTAAATATTTTCG
GTTTTGGGACAGTGGTTGATGGGGCGGTTTACATTTTATGTGTATGTACTTGTTCCTCA
AAACCATTCAATTTTCAAA

>G716 Amino Acid Sequence (domain in AA coordinates: 24-355)
MASVEGDDDFGSSSSRSYQDQLYTELWKVCAGPLVEVPRAQERVFYFPQGHMEQLVASTN
QGINSEEIPIVFDLPPKILCRVLDVTLKAEHETDEVYAQITLQPEEDQSEPTSLDPPIVGP
TKQEFHSFVKILTASDTSHTGGFSVLKHAETCLPSLDMTQATPTQELVTRDLHGFWEFRF
KHIFRQPRRHLLTTGWSTFVSSKRLVAGDAFVFLRGENDLRVGVRRLARHQSTMPTSV
ISSQSMHLGLVATASHAVRTTTFVVFYKPRISQFIVGVNXYMEAIAKHGFSLSGTRFRMRF
EGEESPERIFTGTIVGSDLSSQWPASKWRSLLQVQWDEPTTVQRPDKVSPWEIEPFLATS
PISTPAQQPQSKCKRSRPIEPSVKTPAPPSFLYSLPQSQDSINASLKLFDPSLERISGG
YSSNNSFKPETPPPTNCSYRLFGLDLSNSPAPIPDQKQPMDCGAACQEPITPTSMS
EQKKQQTSRSTKVQMGIAGRAVDLTLKSYDELIDELEEMFEIQQLLARDKWIVVF
TDDEGDMMLAGDDPWNEFCKMAKKIFYSSDEVKMTTKLKISSSLENEEYGNESFENRS
RG*

>G725 (46..1122)
CCTCTTTTCAAGAGAGAAAGAGAGTCAGAGAGAGAGAGAGAGAATGTTCCATGCTAAG
AAACCTTCAAGTATGAATGGTTTCATATGAGAACAGAGCTATGTGCGTTCAAGGCGATTCA
GGCCTTGTCTCTACCACCGACCCCTAAACCGCGTTTGCCTTGGACCGTCGAACCTCCACGAG
CGTTTTGTGGACGCCGTCGCTCAGCTCGGCGGCCCCGACAAAGCGACCCCAAGACGATT
ATGAGAGTTATGGGTGTGAAGGGTCTTACTCTTTACCACCTAAAGAGCCATCTTCAGAAA
TTCAGGCTTGGAAAGCAGCCGCACAAGGAGTACGGAGATCACTCCACAAAGGAAGGTTCA
AGAGCTTCTGCCATGGATATTACAGCGCAACGTAGCTTCTTCTTCTGGCATGATGAGTCGC
AACATGAATGAGATGCAAATGGAAGTGCAGAGAAGGTTGCATGAACAGCTAGAGGTGCAA
AGACATCTGCAACTGAGGATTGAAGCACAAGGAAAGTACATGCAATCTATCTTGGAGAGA
GCTTGCCAAACCTAGCCGGTGAGAACATGGCAGCCGCCACCGCAGCAGCCGCCGTCGGA
GGAGGATACAAGGGTAATCTGGGAAGTTTCGAGTCTTTCAGCAGCGGTGGGCCCCACCTCCT
CATCCTCTTAGTTTCCCGCCGTTTCAAGACCTAAACATCTATGGAAACACAACCGACCAA
GTCCTCGACCATCAAACTTCCATCATCAAAACATAGAGAACCATTTCACGGGTAACAAT
GCTGCAGACACCAACATTTACTTGGGGAAGAAGCGACCTAATCCTAATTTTGGTAACGAT
GTAAGGAAAGGACTATTGATGTGGTCTGATCAAGATCACGATCTTTCGCAAAACCAATCG
ATCGATGATGAGCATAGAATTAGATACAGATGGCTACACATGTCTCCACGGATTGGAT
TCTTTGTGCGAGATCTACGAAAGGAAATCAGGTTTATCAGGTGATGAAGGGAATAATGGT
GGGAAATTACTGGAAGGCCATCGCTAGGAGATCACCATTGAGTCTTATGATGAACCCCT
AATGGTGGATTAATACAAGGAAGAACTCGCCATTGTTGGGTGATACAATTTATTAATTTTT
ATCTATGAGTGATGCATGGGAATGTAAGAACGAGATATATATGTTTTGTCATTGTGAGTT
TGACGTAGGGTTTAGAGAAAA

>G725 Amino Acid Sequence (domain in AA coordinates: 39-87)
MFHAKKPSSMNGSYENRAMCVQGDGLVLTDPKPRLRWTVELHERFVDAVAQLGGPDKA
TPKTIIMRVMGVKGLTLYHLKSHLQKFRLGKQPHKEYGDHSTKEGSRASAMDIQRNVASSS
GMMSRNMNEMQMEVQRRLEHQLVQRHLQLRIEAQKYMQSILERACQTLAGENMAAATA
AAAVGGGYKGNLSSSLSAAVGPPPHPLSFPPFDLNIYGNITDQVLDHNFHHQNIENH
FTGNNAADTNIYLGKKRPNPNFNDVRKGLLMWSDQDHDLSANQSIDDEHRIQIQMATHV
STDLDLSLSEIYERKSLSGDEGNNGKLLERPSRRSPLSPMNPNGGLIQGRNSPFG*

>G727 (43..1977)
CTTCTTCTCCTTCTCTGATCGTTTCGTTTTCTGGACGAGAGAGATGGTAAATCCGGGTAC
GGAAGAGGACCCGATTCCGGGTACTGCTGCTGGTGGGTCAAACCTCCGACCCGTTTCTGCG

AATCTTCGAGTTCTTGTCTGATGATGATCCAACCTGTCTCATGATCTTAGAGAGGATG
CTTATGACTTGTCTCTACAGAGAGCAGAGAGCGCATTGTCTCTGCTTCGGAAGAACAAAG
AATGGTTTTGATATTGTCTATTAGTGATGTTTACATATGCCTGACATGGATGTTTCAAGCTC
CTTGAACACGTTGGTTTAGAGATGGATTTACCTGTTATCAATCTGAATGTTTTGAAACCT
TTGGTTATAGTGATGTCTGCGGATGATTCTGAAGAGCGTTGTGTTGAAAGGAGTGACTCAC
GGTGCAGTTGATTACCTCATCAAACCGGTACGTATTGAGGCTTTGAAGAATATATGGCAA
CATGTGGTGCGGAAGAAGCGTAACGAGTGGAATGTTTCTGAACATTCTGGAGGAAGTATT
GAAGATACTGGCGGTGACAGGGACAGGCAGCAGCAGCATAGGGAGGATGCTGATAACAAC
TCGTCTTCAGTTAATGAAGGGAACGGGAGGAGCTCGAGGAAGCGGAAGGAAGAGGAAGTA
GATGATCAAGGGGATGATAAGGAAGACTCATCGAGTTTAAAGAAACCACGCGTGGTTTGG
TCTGTTGAATTGCATCAGCAGTTTGTGCTGTGTAATCAGCTAGGCGTTGACAGTGAG
TTAAAACTTGCTTGCTTATGCATTGTGTGTGTCGATTGGTAACATTGTGGAATCCAG
AAGTATCGGATATATCTGAGACGGCTTGGAGGAGTATCGCAACACCAAGGAAATATGAAC
CATTTCGTTTATGACTGGTCAAGATCAGAGTTTTGGACCTCTTTCTTCGTTGAATGGATTT
GATCTTCAATCTTTAGCTGTTACTGGTCAGCTCCCTCCTCAGAGCCTTGACACAGCTTCAA
GCAGCTGGTCTTGGCCGGCTTACACTCGCTAAACCAGGGATGTCGGTTTCTCCCTTGTA
GATCAGAGAAGCATCTTCAACTTTGAAAACCCAAAAATAAGATTTGGAGACGACATGGT
CAGACGATGAACAATGGAAATTTGCTTCATGGTGTCCCAACGGGTAGTCACATGCGTCTG
CGTCTGGACAGAATGTTTCAAGCAGCGGAATGATGTTGCCAGTAGCAGACCAGCTACCT
CGAGGAGGACCATCGATGCTACCATCCCTCGGGCAACAGCCGATATTGTCAAGCAGCGTT
TCAAGAAGAAGCGATCTCACTGGTGCCTGGCGGTTAGAAACAGTATCCCCGAGACCAAC
AGCAGAGTGTACCAACTACTCACTCGGTCTTCAATAACTTCCCCGCGGATCTACCTCGC
AGCAGCTTCCCGTTGGCAAGTGCCCCAGGGATTTAGTTCCAGTATCAGTTTCTTACCAA
GAAGAGGTCAACAGCTCGGATGCAAAAGGAGGTTTATCAGCTGCTACTGCTGGATTGGT
AACCAAGCTACGACATATTTAACGATTTTCCGAGCACCAACAGCACAAAGAACATC
AGCAATAAATAACGATTGGGATCTGCGGAATATGGGATTGGTCTTCAAGTTCCAATCAG
GACGCAGCAACTGCAACCGCAACCGCAGCATTTTCCACTTCGGAAGCATACTCTTCGTCT
TCTACGCAGAGAAAAGACGGGAACCGGACGCAACAGTTGTGGGTGAGCATGGGCAGAAC
CTGAGTCAACGAGCCGGAATCTGTATCATCTGAACACGTTTATGAGCGGTGTTTCA
GTCAGAGTGAACTCAGAAAGAGTGGCGGAGACAGTGACTTGTCTCCAGCAAATACATTG
TTTACAGAGCAGTATAATCAAGAAGATCTGATGAGCGCATTTCTCAAACAGGTTTGATTA
TTACTCGAATACAGTGCACTCTAAAC

>G727 Amino Acid Sequence (domain in AA coordinates: 226-269)
MVNPGHGRGPDSTAGGSNSDPFANLRVLVDDDPCLMILRMLMTCLYREQRAHCL
CFGRKTNGFDIVISDVHMPDMDGFKLLEHVGLEMDLPVINLNLKPLVIVMSADDSKSVV
LKGVTGHGAVDYLIKPVRIEALKNIWQHVVRRKRNENNVSEHSGSIEDTGDRDRQQQHR
EDADNNSSVNEGNRSSRRKKEEVDDQDDKEDSSSLKPRVWVSVLHQQFVAAVNQ
LGVDSELKTCLLMHLCVSIGNIVEFQKYRIYLRLGGVSOHQGNMNSFMTGQDQSFGLP
SSLNGFDLQSLAVTGQLPPQSLAQLQAAGLGRPTLAKPGMSVSPLVDQRSIFNFENPKIR
FGDGHGQTMNNGNLLHGVPTGSHMRLRPGQNVQSSGMMLPVADQLPRGGPSMLPSLGQQP
ILSSSVSRRLTGALAVRNSIPETNSRVLPPTHVSFNNFPADLPRSSFPLASAPGISVP
VSVSYQEEVNSSDAKGGSSAATAGFGNPSYDIFNDFPQHQQHNKNI SNKLNWDLRNMG
VFSSNQDAATATATAAFSTSEAYSSSSQQRKRETDATVVEHGNLQSPSRNLYHLNHV
FMDGGSVRVKSERVAETVTCPPANTLFHEQYNQEDLMSAFLKQV*

>G740 (25..924)

CTTCTTCAACTTTTTTTTTTAAACGATGGCTTCAGAGGATCAATCGGCGGCGAGATCTACC
GGGAAGGTGAAGTGGTTCAACGCTTCTAAAGGCTATGGTTTCATTACTCTGACGATGGC
AGCGTAGAGCTTTTGTTCATCAATCTTCAATTGTCTCCGAAGGTTACCGGAGTTTAAAC
GTCGGCGATGCGGTTGAGTTCGCTATTACTCAGGAAGCGACGGTAAGACTAAAGCCGTC
AATGTTACTGCTCCTGGTGGTGGTTCTCTCAAGAAGGAGAATAACTCTCGTGGTAACGGT
GCTAGGCGCGGCGGCGGTGGAAGCGGTTGCTACAATTGCGGTGAGTTAGGTATATCTCT
AAAGATTGTGGTATTGGTGGCGGCGGAGGTGGTGAACGTAGATCTAGAGGAGGAGAA
GGTTGTTCAAAATGTGGTGATACTGTCACTTCGCTAGGGATTGTACTTCAGCTGGAAC
GGTGACCAACGTGGAGCCACCAAGGTGGAACGATGGTTGCTACACTTGCCTGATGTT
GGTCACGTGGCTAGGGATTGTACTCAGAAATCAGTTGGAACGGAGACCAACGTGGAGCG
GTCAAAGGTGGAACGATGGTTGCTACACTTGTGGTGTGTTGGTCACTTTGCTAGGGAT
TGTAATCAGAAGGTTGCTGCCGGAACGTCAGAAGCGGTGGTGGTAGTGGAACCTTGT

TATTCATGCGGTGGAGTTGGTCACATTGCAAGAGATTGTGCGACTAAGAGACAGCCTTCT
CTGGGGTGTGTACCAGTGTGGTGGTTCTGGTCACTTGGCTCGTGATTGTGACCAGAGAGGA
AGCGGTGGAGGAGTAATGATAATGCGTGCTACAAGTGTGGTAAGGAAGGTCACCTTTGCA
AGGGAATGTTCTTCTGTAGCTTAATCGATTTCTAATCAACAAAAACAAAAACAAGAAT
GAAATTGAATCGAGTTATATAGTTTGGTATATATTACTCTTCGTTTTTCATTTATCTTTTT
TTTTGTTGTTGATGGGAATGAAATTGCCTGGTCCTTTTGGTGTGTTTTTGAGCTTTTATT
ATTATACAGAGTGATCCCTTTTTTGTATAACTATTACAAGTTTTTAGCTTTATTTGATA
TGGATGCTCTCTCCTTTTCTTCTATCTGTTTCTGGAATTTTGACCTCATCATATTACTT
ATGTCATCCAAA

>G740 Amino Acid Sequence (domain in AA coordinates: 24-42, 232-268)

MASEDQSAARSTGKVNWFNASKGYGFITPDDGSVELFVHQSSIVSEGYRSLTVGDAVEFA
ITQGSDDGKTKAVNVITAPGGGSLKKNNSRNGARRGGGSGCYNCGELGHISKDCGIGGG
GGGGERRSRGEGCYNCGDTGHFARDCTSAGNGDQRGATKGGNDGCTCGDVGHVARDCT
QKSVGNGDQRGAVKGGNDGCTCGDVGHFARDCTQKVAAGNVRSGGGSGTCYSCGGVGH
IARDCATKRQPSRGYQCGSGHLARDCDQRSGGGGNDNACYKCGKEGHFAECSSVA*
>G770 (119..1069)

CCTTCCTCTATATAAGGAAGTTCATTTTCATTTGGAGAGGACACGCTGACAAGCTGACTCT
AGCAGATCTGGTACCGTCGACGGTTCTTGGATTGGAGTAACTAAAGATCATATAAAAT
GGAACAAGGAGATCATCAGCAGCATAAGAAAGAAGAAGACTTTGCCACCGGGTTTCAG
ATTTTCATCCGACCGATGAGGAGCTAATCTCATATTACTTGGTTAATAAGATTGCCGATCA
AACTTCACCGGGAAGCAATCGCTGACGTTGATCTTAACAAGTCCGAGCCATGGGAGCT
TCCTGAGAAGGCGAAAATGGGAGGAAAAGAATGGTACTTTTTTAGCCTCCGGGACCGGAA
GTACCCGACGGGAGTGAGGACGAATAGGGCGACGAATACAGGATATTGGAACACACAGG
AAAAGACAAAGAGATATTCAATAGCACAACTCGGAGTTGGTCCGGGATGAAGAAGACTTT
GGTCTTTTACAGAGGACGAGCTCCTCGTGGGGAGAAGACTTGTGGGTCTATGCATGAGTA
TCGACTTCACCTCAAGTCTCATATAGAACCTCCAAGCAAGACGAGTGGGTAGTGTGTAG
AGTGTTCAAGAAAACAGAAGCAACCAAGAAATACATAAGCACCAGTAGCAGCAGCACAAG
TCATACCACAACAACCAACAAGAGCCTCAATACTATCAACCAACAATAATCCTAA
TTACTCATCAGACCTCCTTCAACTCCACCGCATCTACAACCAACCCGAGCCTCAATAT
TAACCAATCCCTCATGGCAAACGCCGTTACCTAGCTGAGCTCTCAAGAGTCTTCCGTGC
CTTACAAGCACCACCATGGACTCTTCTCATCAGCAGCTAATGAATACACCCACATGCC
TGTTCTCAGGGCTCAACCTCAACCTTGGCGGTGCACTGGTCCAGCCGCCTCCTGTTGTGTC
TCTTGAGGATGTTGCCCGGTTAGTGCTTCGTACAATGGCGAAAACGGGTTTGGAATGT
GGAGATGAGCCAGTGATGGACTTGGATGGATACTGGCCATCTTATTGATTGGTAATTGT
CAGTTTAAGTTATGGTTTTTATATTGTTTCCATTTACTTGTGGTAAAACGATTTTGTT
GTTCTTGCGAACCTCTAGACAGGCCTCGTACCGGATCCTCTAGCTAGAGCTTTCGTTTCG
TATCATCGGTTTC

>G770 Amino Acid Sequence (domain in AA coordinates: 19-162)

MEQGDHQHKKKEEALPPGFRFHTDEELISYYLVNKIADQNFTGKAIADVDLNKSEPWE
LPEKAKMGGKEWYFFSLRDRKYPTGVRTNRAATNTGYWKTGKDKKEIFNSTTSELVGMKKT
LVFYRGRAPRGEKTCWVMHEYRLHSKSSYRTSKQDEWVVCRVFKKTEATKKYISTSSST
SHHHNNHTRASILSTNNNNPNYSSDLLQLPPLHQLPHPSLNLNQLSLMANAVHLAELSRVFR
ASTSTTMDSSHQQLMNYTHMPVSGNLNLGGLVQPPPVVSLDVDAAVSASYNGENGFN
VEMSQCMDLDGYWPSY*

>G858 (99..869)

CATAATCTCTTCTCTATATCTCTTCTCTTCTTTTACCCTGTTTTTTTTTTCATTC
CACAGAGCCCAGGTTGATTGATTTTGTATTTCAGAGATATGGGGAGAGGAAGGATTGAGA
TTAAGAAGATTGAGAATATCAACAGTCGTCAAGTCACTTTCTCTAAGAGACGAAACGGTT
TGATCAAGAAGGCTAAAGAGCTTTCGATTCTCTGTGACGCCGAGGTTGCTCTTATCATCT
TCTCCAGCACCAGCAAGATTTACGATTCTCCAGCGTCTGTATGGAGCAAAATCTTTCTA
GATATGATACACTACTGCGTCCACTGAGCATAAACAACAAGAGAACACCAACTTCTAA
TTTGTGCTTCACATGGAAATGAAGCTGTGTTGCGAAATGATGATTCTATGAAGGGGGAAC
TTGAAAGATTACAGCTTGCAATTGAGAGACTTAAGGGTAAGGAGCTTGAAGGTATGAGTT
TCCCGGATCTTATTTCTCTTGAACACAGTTGAACGAGAGCTTGATAGTGTCAAGGATC
AAAAGACACAAATCCTGCTCAACCAGATTGAGAGATCCAGGATACAGGAGAAAAAAGCAT
TGGAAGAAAACCAATCTTGCAGCAACAGGTTGAGATGTTGGGGAGAGGTTGAGGACCA
AAGTGTGTAATGAAAGGCCTCAAGATTCTAGCCAGAACCGATCCCGAGAGCTCTTCAT

CAGAAGAGGATGAGAATGACAACGAGGAGCACCATTCCGACACTTCCTTGACAGTTGGGGT
TGTCGTCGACGGGGTATTGCACAAAGAGAAAGAAGCCGAAGATCGAACTGGTCTGCGATA
ACTCTGGGAGTCAAGTGGCTTCTGATTGATGGAATCGATTATTTTCTAATTCTGGTTGT
TTAGGGGTCTCTATGTGTCTTCTGTTTCTGGCTGTTCTTTGCTTTATTTTCATCTCAAG
TAGAGTTTTCTTAATGTTTAGGTGGAACATTTTTCCATAATCAAGAAGGGATTGATCAA
TCAATAACATTAGATTTTCTTAGTTAAAGACTTAAAGTTGCCACACACCACACCATATG
TGATTATGATGAATTTACATTTTATAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA

>G858 Amino Acid Sequence (domain in AA coordinates: 2-57)

MGRGRIEIKKIENINSRQVTFSKRRNGLIKKAKELSILCDAEVALIIFSSTGKIYDFSSV
CMEQILSRGYGTTASTEHKQREHQLLICASHGNEAVLRNDDSMKGELERLQLAIERLKG
KELEGMSFPDLISLENQLNESLHSVVDQKTQILLNQIERSRIQEKKALEENQILRKQVEM
LGRGSGPKVNLNERPDQSSPEADPESSSSEEDENDNEEHSDTSLQLGLSSTGYCTKRKKP
KIELVCDNSGSQVAD*

>G865 (282..920)

ATCCCCACTTGTGTTTCATCACCAAGCCAAGCTCCATGTCCTAGTCACTCCACAGATTCC
CTATCATCATCAATTCGTTTCAAACCTAGTTCCTTTCAAAGTCTTGACATATATACACA
CACACCTATTATTCTCTTGGTGTGTTGTGTGTACATATACGTGTGAGTACATACTTTG
TTGTAAAAGTGGATCGGAGGTATGGAAAGGGACCGGTTCCACCGGAAACATCGGCGGCGG
CGGATGATAATTCGTCTTGAACGAGACTGATGTCACCGCCATGGTCTCCGCTCTCAGCC
GTGTCATAGAGAATCCGACAGACCCGCGGTCAAACAAGAGCTTGATAAATCGGATCAAC
ATCAACCAGACCAAGATCAACCAAGAAGAAGACACTATAGAGCGTAAGGCAGAGACCAT
GGGGTAAATGGGCGGCAGAAATCCGCGATCCAAAGAAAGCAGCCCGTGTCTGGCTCGGGA
CTTTCGAGACGGCAGAGGAAGCTGCTTTAGCCTATGACCGAGCTGCCCTCAAATTCAAAG
GCACCAAGGCTAACTGAACTTCCCTGAACGGGTCCAAGGCCCTACTACCACCACAACCA
TTTCTCATGCACCAAGAGGAGTTAGTGAATCCATGAACTCACCTCCTCCTCGACCTGGTC
CACCTTCAACTACTACTACTTTCGTGGCCAATGACTTATAACCAGGACATACTTCAATACG
CTCAGTTGCTTACGAGTAACAATGAGGTTGATTTATCATACTACAGTCGACTCTCTTCA
GTCAACCTTTTTCAACGCCTTCTTCATCTTCTTCTCTCCCAACAGACGCAGCAACAGC
AGCTACAACAACAACAACAGCAGCGTGAAGAAGAAGAAGAAATTATGGTTACAAATTATT
ATAACTACCCAAGAGAATAATCTAATTATTATTGTTGGTCAATCAGTTTTATAAATAGC
TATCATAGTTTCATTTTTGGTTTCCGTAACTTTGTTGCATGGAAAATATGAATGAACGA
GGGACATGTGTAACAATTTGTTTGTGTTTCGTAAATGTTAGTTGATTTGGATTGCTGA
AGTTTGATTTTTCTGAGCATAAATCATTGACGGTCAAAAAAAAAA

>G865 Amino Acid Sequence (domain in AA coordinates: 36-103)

MVSALSRVIENPTDPPVKQELDKSDQHQPDDQPRRRHYRQVRQRPWGKWAABEIRDPKKA
ARVWLGTFFETAEAAALAYDRAALKFKGTKAKLNFPERVQGPTTTTTISHAPRGVSESMNS
PPRPGPSTTTTTSWPMTYNQDILQYQLLTSNNEVDLSYYTSLFSPQPFSTPSSSSSSS
QQTQQQQLQQQQQQREEEKNYGYNYNYPRE*

>G872 (59..646)

CCGGAACAGAAATCCAATTCAACCAAACCGAATCGAACCGAACCGGAGTTTTTATCCAAT
GGTGAAGCAAGCGATGAAGGAAGAGGAGAAGAAGAGAAACACGCGGATGCAGTCAAAGTA
CAAAGGAGTGAGGAAGAGGAAATGGGGAAATGGGTATCGGAGATCAGACTTCCACACAG
CAGAGAACGAATTTGGTTAGGCTCTTACGACACTCCCAGAAAGGCGGCGGTGCTTTTCA
CGCCGCTCAATTTTGTCTCCGCGGCGGCGATGCTAATTTCAATTTCCCTAATAATCCACC
GTCGATCTCCGTAGAAAAGTCGTTGACGCTCCGAGATTTCAGGAAGCTGCTGTAGATT
CGCTAACACATTCCAAGACATTGTCAAGGGAGAAGAAGAATCGGGTTTAGTACCCGGATC
CGAGATCCGACCAGAGTCTCCTTCTACATCTGCATCTGTTGCTACATCGACGGTGGATTA
TGATTTTTTCGTTTTTGGATTGCTTCCGATGAATTTCCGGTTTGATTCTTCTCCGACGA
CTTCTCTGGCTTCTCCGTGGTGATCGATTTACAGAGATTTTACCCATCGAAGATTACGG
AGGAGAGAGTTTTATTAGATGAATCTTTGATTCTTTGGGATTTTTGAATTTCCAAACATAA
TATTTTTTTAGAGCGAACTGTGAGATTTTCTTGGAGTCATGGAGAAATCTGGAGATTTT
TTGTAACACGGAGCTCAATGACCCGGGAATTTCTTTGTTTCGGATCCGAATTTGATGT
GGATCATATTACACCTATATTTTTTTCATTTTTTTGTTGTAAAGAAAAATCGGATAAGAT
TCTAGTAATAAATGTTAAAGTCCATTTTCATTAATAAAAAAAAAAAAAAAAAA

>G872 Amino Acid Sequence (domain in AA coordinates: 18-85)

MVKQAMKEEEKRNTAMQSKYKGVKRKWKVWSEIRLPHSRERIWLGSYDTPEKAARAF
DAAQFCLRGGDANFNFPNNPPSISVEKSLTPPEIQEAAARFANTFQDIVKGEEESGLVPG

SEIRPESPSTASVATSTVDYDFSFLDLLPMNFGFDSFSDDFSGFSGGDRFTEILPIEDY
GGESLLDESLILWDF*

>G904 (1..1005)

atggaatctctcatcaatcccagccatggcggaggaaactacgattctcactcttctctct
ctcgatagctctcaaaccaagcgtactagtcattctcattctcctcatgactcttctc
atctccgtttccatttgcttctcctccgctgtctcaatcgctgtagccaccgctccgtt
ctccctcttccatcttctcttccgtcgcaaccgtaacttccgattcccgaagattctct
ggacatcgagctctctcccgaacagaacggctcctccgtgcttgattcgcttccgattttc
aaattctcctccgtcactcgccgatctagctccatgaattccggagattgcgcgcttgg
ttgtcgaaattcgaaacggaggatcagctccgtcttcttctctctgttgctcacgctttt
cacgccgattgtatctggtatgtcttaaccagactgtcctctctgtcgctct
cctctcttccgttcagaatctgatctcatgaagtctctcgccgctcgctcggtcaaacaac
ggcggaggagaaaacagcttccgtctcgaaatcggatccatcagccgtcgctcgtaaca
ccgattccagaatccgttgagcagcatcgaacttactcaatcggttcgttcgattacata
gtagacgagctagattcagaaatctcagagtcgaatttcaaccgtggaaaacaggaagac
gcgactacaacaactgccacagcaacggcggttacgactaatccgacgtcggttgaagct
agtttagcggcggtataggttaacgatggttctagaagctggctcaaggattacgttgac
agactctcacgaggtatctcgctcggtgcaatgtcggttagaagctctggtagattttt
actgggagtagtcgtcgagtgaggaattgacggtgatggatttagaagcgaatcatgcc
ggagaagagataagtgagcttttccggtggctctcaggggtgta

>G904 Amino Acid Sequence (domain in AA coordinates: 117-158)

MESLINPSHGGGNYDSHSSSLSLKPSVLVIIILLMTLLISVSICFLLRCLNRCSHRSV
LPLSSSSSVATVTSRRRFSGHRVSPETERSSVLDLPIFKFSSVTRSSSMNSGDCAVC
LSKFEPEQLRLPLCCHAFHADCIDIWLVSNQTCPLCRSPLFASESLMKSLAVVGSNN
GGGENSFRLEIGSISRRTPIESVEQHRYSIGSFYIVDDVDSEISESNFNRGKQED
ATTTTATATAVTTNPTSFASLAADIGNDGSRSLWKDYVDRLSRGISSRAMSFRSSGRFF
TGSSRRSEELTVMDLEANHAGEEISELFRWLSGV*

>G910 (1..1071)

ATGTTATGTATAATAATAATTGAGAATATGGAAAGAGTATGTGAGTTTGTAAAGCGTAT
AGAGCAGTGGTTTATTGTATAGCTGATACAGCAAATCTTGTTTAACATGTGATGCAAAG
GTTTCATTACGCTAATTCACCTCTCGGGACGGCATTACGTACGGTTTTATGTGATTCTGGT
AAGAATCAGCCTTGTGTTGTCCGATGTTTTGACCATAAAATGTTCTTTGCCATGGATGT
AATGATAAGTTTCATGGTGGTGGCTCTTCTGAGCATCGTAGAAGGGATTGAGGTGTTAT
ACGGGTTGTCTCTCTGCTAAAGATTTGCGGGTTATGTGGGGTTTTGAGTTATGGATGAC
GATGATGATGTTTCGTTAGAGCAATCTTTTGAATGGTTAAACCTAAGGTGCAAAGAGAA
GGTGGTTTTATCTTGAACAGATTTCTTGAATTGGAGAAGGTTTCAGCTCAGGGAAGAGAAT
GGTAGTTCTTCTTGACAGAACGAGGTGATCCATCTCCATTGGAGCTTCCTAAGAAACCC
GAAGAACAGTTAATCGATCTTCCGACAGACCGGAAAAGAGCTGGTTGTTGATTTTTCACAC
TTGTCTCATCTTCCACACTTGGTGATTCCTTTGGGAATGCAAAAGTCCATACAATAAG
AACAAATCAGTTGTGGCATCAAATATACAAGACATTGGAGTATGTGAAGATACAATCTGC
AGTGACGATGACTTCCAAATACCTGACATTGATCTCACTTTCCGGAACTTTGAAGAGCAA
TTTGGAGCTGATCCTGAGCCAATTGCAGATAGTAACAACGTGTTCTTTGTTTCTTCCCTT
GACAAATCAGATGAGATGAAGACATTTTCTTCTTCAATAATCCCATATTTGCACCT
AAACCAGCTTTCATCACTATCTCATTCTCAAGCAGTGAAACCGATAACCCCTTATAGTCAC
TCAGAGGAAGTAATCTCATTTTGTCCCTCCCTCTCTAACAATACACGTCAAAAGGTCATC
ACAAGGCTCAAGGAGAAGAAGAGAGCAAGAGTGGAGGAGAAAAAGCTTAA

>G910 Amino Acid Sequence (domain in AA coordinates: 14-37, 77-103)

MLCIIIIENMERVCEFCAYRAVVYCIADTANLCLTCDKAVHSANSLSGRHLRTVLCDSG
KNQPCVVRCFDHKMFCLCHGCNDKFHGGSSSEHRRRLRCYTGCPPAKDFAVMWGFRVMD
DDDSVLEQSFMRVMPKVQREGGFLEQILELEKVQLREENGSSSLTERGDPSPLELPKP
EEQLIDLPTGKELVDFSHLSSSSSTLGDSECKSPYNKNQLWHQNIQDIGVEDTIC
SDDDFQIPDIDLTFRNFEEQFGADPEPIADSNVFFVSSLDKSHEMKTFSSSFNNPIFAP
KPASSTISFSSSETDNPYSHSEEVISFCPSLSNNTRQKVIITRLKEKKRARVEEKA*

>G912 (20..694)

CATCTTATCCAAAGAAAAATGAATCCATTTTACTCTACATTTCCAGACTCGTTTCTCTC
AATCTCCGATCATAGATCTCCGGTTTCAGACAGTAGTGAGTTTACCAAAGTTAGCTTC
AAGTTGTCCAAAGAACGAGCTGGGAGGAAGAAGTTTCGTGAGACACGTATCCGATTTA

CAGAGGAGTTCGTCAGAGGAATTCTGGTAAATGGGTTTGTGAAGTTAGAGAGCCTAATAA
GAAATCTAGGATTTGGTTAGGTACTTTTCCGACGGTTGAAATGGCTGCTCGTGCTCATGA
TGTTGCTGCTTTAGCTCTTCGTGGTCTGCTCTGCTTGTCTCAATTTTCGCTGATTCTGCTTG
GCGGCTTCGTATTCTTGAGACTACTTGTCTAAGGAGATTAGAAAGCTGCGTCTGAAGC
TGCAATGGCGTTTTCAGAATGAGACTACGACGGAGGGATCTAAACTGCGGCGGAGGCAGA
GGAGGCGGACGGGGAGGGGTGAGGGAGGGGAGAGGGGCGAGGAGCAGAATGGTGG
TGTGTTTTATATGGATGATGAGGCGCTTTTGGGGATGCCCACTTTTTTGAGAATATGGC
GGAGGGGATGCTTTTGGCGCCGCCGAAGTTGGCTGGAATCATAACGACTTTGACGGAGT
GGGTGACGTGTCACTCTGGAGTTTGTACGAGTAATTTTTGGCTCTTTTCTGGATAATA
AGTT

>G912 Amino Acid Sequence (domain in AA coordinates:51-118)
MNPFFYSTFPDSFLSISDHRSPVSDSSECSPKLASSCPKKRAGRKKFRETRHPIYRGVRQR
NSGKWVCEVREPNKKSRIWLGTFTVEMAARAHDAALALRGRSACLNFAWSAWRLRIPE
TTCPKIEIQKAASEAAMAFQNETTTEGSKTAAEAEEAAGEGVREGERRAEEQNGGVFYMD
EALLGMPNFFENMAEGMLLPPEVGWNHNDGVDGVDLSLWSFDE*

>G920 (114..1154)

AAAAAATCTATTTCTTCTCTTTCCACTATATTACAACATTTCTTCATTCTCAAATCATC
ATACTAAAAACCTAAAAAAGTTACATATTATTGTATCTTTGTGAGAAAAAATGGATT
CGAATAGTAACAACACGAAATCCATAAAGAGAAAAGTTGTGACCAACTTTGTGAAGGCT
ATGAATTCGCTACTCAGCTTCAGCTTCTCCTTTCTCATCAACACTCTAACAGTACCACA
TCGATGAGACCCGCTCTTGTTCGCGGTTCGGGTTTTCGGGTTCAGTTTCCGGTGGTCCAGATCCCGTTG
ATGAGCTCATGTCTAAGATCTTGGGATCTTTCCATAAACTATATCGGTTCTTGATTCTT
TTGATCCCGTCGCGCTCTCTGTCCCATCGCCGTCGAGGGTTCATGGAATGCTTCATGTG
GGGATGATTCGCGCACTCCGGTGAGTTGCAACGGTGGAGATTCCGGTGAGAGTAAGAAGA
AGAGATTAGGGGTTGGTAAGGGTAAAAGAGGATGCTACACTAGAAAGACGAGATCACATA
CAAGGATCGTGGAAGCTAAAAGTTCTGAAGACAGATATGCTTGGAGGAAATATGGACAAA
AGGAGATTCTTAATACCACATTTCCCAAGAAAGTTACTTTAGATGCACACACAAGCCAACGC
AAGGATGCAAAGCAACAAAGCAAGTTTCAGAAAACAGGATCAAGATTCTGAGATGTTCAAA
TCACATACATTGGCTACCAACATGCACACTGCCAATGACCAACGCACGCGAAGACCGAGC
CTTTTGATCAAGAAATCATTATGGATTTCGAAAAGACATTGGCTGCTAGCACTGCTCAGA
ACCATGTCAATGCTATGGTGCAAGAGCAAGAGAACACACCAGCAGTGTGACAGCAATAG
ACGCAAGCATGGTTAAGGAGGAACAAAATAACAATGGTGATCAGAGTAAAGATTATTATG
AGGGCTCTTCGACAGGTGAGGACTTGTCAATTGGTTTGGCAAGAGACGATGATGTTTGATG
ATCATCAAAATCACTACTATTGTGGTGAAACAGTACTACTTCTCATCAATTTGGTTTCA
TCGACAACGATGATCAGTTTTCTCTCTTCTTCGACTCATATTGTGCTGATTATGAAAGAA
CAAGTGCTATGTGAACATCCAAATCTGGAATGATGAATCAGCACTAGGTCTTCTCTTTGA
GTATGTCTAGTTTAATGTAAATATTTTGTGTATGTTTGATAAAAACACCATATATACTT
CTCTTTTACACCAAAAAAAAAAAAAAAAAAAAAA

>G920 Amino Acid Sequence (domain in AA coordinates: 152-211)
MDSNSNNTKSIKRKVVQDLVEGYEFATQLQLLLSHQHSNQYHIDETRLVSGSGSVSGGPD
PVDELMKILGSHFKTISVLDSFDPVAVSVPIAVEGSWNASCGDDSATPVSCNGGDSGES
KKKRLGVGKGKRGCYTRKTRSHTRIVEAKSSEDRYAWRKYGQKEILNTTFPRSYFRCTHK
PTQGCKATKQVQKQDQDSEMFQITYIGYHTCTANDQTHAKTEPFQEIIMDSEKTLAAS
AQNHVNAMVQEQENNTSSVTAIDAGMVKEEQNNNGDQSKDYEGSSSTGEDLSLVWQETMM
FDDHQNHYYCGETSTTSHQFGFIDNDQFSSFFDSYCADYERTSAM*

>G939 (9..1565)

CAGATTCTATGGATATGTATAACAACAATATAGGGATGTTCCGGAGTTTAGTTTGTAGCT
CGGCGCCTCCATTTACAGAGGGACATATGTGTTCTGATTTCGCATACGGCTTTGTGCGATG
ATCTGAGTAGTGATGAGGAAATGGAAATAGAGGAGCTTGAGAAGAAGATCTGGAGAGACA
AGCAGCGTTTTAAAGCGCTCAAGGAAATGGCGAAGAACGGTCTAGGAACAAGATTGTTGT
TGAAGCAGCAACATGATGATTTTCCAGAGCACTCTAGTAAGAGAACCATGTACAAGGCAC
AAGATGGGATCTTGAAGTACATGTCGAAGACAATGGAGCGATATAAAGCTCAAGGTTTGT
TTTATGGGATTTGTTTAGAGAATGGGAAAACGGTAGCGGGATCTTCTGATAATCTCCGTG
AATGGTGGAAAGACAAAGTGAGGTTTGATAGGAACGGCCAGCTGCTATAATCAAGCACC
AAAGGGATATCAATCTTCTGATGGAAGTGATTCAGGGTCTGAGGTTGGGGATTCTACCG
CACAGAAGTTGCTTGAGCTTCAAGATACTACTCTTGAGCTCTGTTATCGGCTCTGTTTC
CTCACTGCAACCTCTCAGAGGCGGTTTCCGTTGGAGAAAGCGGTGACACCGCCATGGT

GGCCAACGGGGAAAGAAGATTGGTGGGATCAACTGTCTTTACCCGTTGATTTTCGAGGTG
TTCCGCCACCTTACAAGAAGCCTCATGATCTCAAGAAGCTGTGGAAAATTGGTGTTTTGA
TTGGTGTAAATCAGACATATGGCTTCTGACATTAGCAACATACCCAATCTCGTGAGACGGT
CTAGAAGTTTGCAGGAGAAAATGACGTCAAGAGAAGGCGCTTTATGGCTCGCTGCTCTTT
ACCGAGAAAAGGCTATTGTTGATCAAATAGCCATGTCTAGAGAAAACAACAACACTTCTA
ACTTTCTTGTTCCTGCAACCGGTGGAGACCCAGATGTTTTGTTTCTGAATCTACAGACT
ATGATGTTGAACTGATTGGTGGCACTCATCGGACCAATCAGCAGTATCTGAATTTGAAA
ACAACACTAACTGTGTTTACAAGAGAAAGTTTGAAGAAGATTTTGGGATGCCAATGCATC
CAACACTCCTAACATGTGAGAACAGTCTCTGTCTTATAGCCAACCACATATGGGATTTCT
TTGACAGGAACCTTAAGAGAGAATCACCAAATGACTTGTCTTATAAAGTCACTTCCTTCT
ACCAACCAACTAAACCCCTATGGTATGACGGGTTTAAATGGTTCCTTGTCCGATTATAACG
GGATGCAGCAGCAGGTTTCAAGACCACTTAAATCATCCCAACGATCTCTACA
GACCAAAAGCTCCACAAAGAGGCAACGATGACTTGGTTGAGGATTTGAATCCTTCTCCTT
CGACGCTGAATCAGAATCTTGGTTTAGTCTTACCTACTGACTTCAATGGAGGTGAGGAAA
CAGTAGGAACAGAGAACAATCTGCATAATCAAGGGCAAGAGTTGCCACATCTTGGATTC
AGTAAAGAAAGCTTCAAGATTTCTTTTATGTTTTCTAGTCTTTATAGCTTTGTCTCTT
GCTTATTCTCTCATTAAACACAGTTTTTGTATCTCTCCATTTATAGCCCATGTAGCAATG
GAGAAGATTAGGTTTCATAATAAGTTAATAACCAAAATTCAAA

>G939 Amino Acid Sequence (domain in AA coordinates: 97-106)
MDMYNNNIGMFRSLVCSSAPPFTTEGHMCSDSHTALCDDLSSDEEMEIEELEKKIWRDKQR
LKRLKEMAKNGLGTRLRLKQHQHDFPEHSSKRTMYKAQDQILKYMSKTMERYKAQGFVYG
IVLENGKTVAGSSDNLREWWKDKVRFDRNGPAIIKHQRDINLSDGSDSGSEVGDSTAQK
LLELQDTTLGALLSALFPHCNPPQRRFPLEKGVTPPWPTGKEDWWDQLSLPVDFRGVPP
PYKKPHDLKKLWKIGVLIGVIRHMASDISNIPNLVRRSRSLQEKMTSREGALWLAALYRE
KAIVDQIAMSRNNNTSNFLVPATGGDPDVLFPSTDYDVELIGGTHRTNQYPEFENNY
NCVYKRKFEEDEFGMPMHPTLLTCENSLCPYSQPHMGFLDRNLRENHQMTCPIKVTSTFYQP
TKPYGMTGLMVPDPCDYNMGQQVQSFQDQFNHPNDLYRPAQRGNDDLVEDLNPSPTL
NQNLGLVLPDFFNGEETVGTENNLHNQGELEPTSWIQ*

>G963 (1..897)

ATGAGTTTGCTCCAGGATTCAGGTTTCATCCCACTGATGAAGAAGTGGTGGCTTACTAT
CTTGATAGGAAGGTCAACGGCCAAGCCATTGAGCTCGAGATCATCCAGAGTTGATCTT
TATAAATGCGAGCCATGGGACTTGCCTGAAAAGTCATTTTGCCTGGGAAACGACATGGAA
TGGTACTTTTACAGCACAAGGGATAAGAAGTATCCAAATGGCTCTAGGACGAACCGTGCG
ACCCGAGCGGGTTACTGGAAGGCCACGGGGAAAGATCGTACAGTAGAATCAAAGAAGATG
AAGATGGGAATGAAGAAGACACTGGTTTATTATAGAGGAAGGGCTCCTCATGGCCTTCGT
ACTAATTGGGTTCATGATGAATATCGTCTCAGCACGCTCCTTCTCCTCCTTGAAGGAG
TCGTATGCATTGTGCGAGTGTTTAAGAAGAACATACAAATTCCAAAGAGAAAAGGGGAA
GAAGAAGAAGCAGAAGAAGAGAGCACTAGTGTAGGAAAAGAAGAGGAAGAAGAAAAGGAG
AAGAAGTGGAGAAAATGTGATGGTAATTATATTGAAGACGAGAGCTTGAAGAAGCATCC
GCGGAGACATCTTCATCAGAGCTAACTCAAGGGGTCTTTTAGACGAAGCAAACAGCTCA
TCCATATTTGCTCTTCATTTCTCATCTTCTCTTCTGGACGATCATGATCATCTTTCTCA
AACTATTCTCATCAGCTTCCATATCATCTCTCTTCAACTCCAAGATTTCCCTCAACTT
TCTATGAACGAAGCAGAGATTATGTCAATCCAACAAGACTTCAATGCAGAGACTCTATG
AACGGGACACTTGACGAAATCTTCTCTTCTTCCGCCACTTTCCCGCTTCCCTTGA

>G963 Amino Acid Sequence (domain in AA coordinates: TBD)
MSLPPGFRFHPTEELVAYYLDKRVNGQAIIELEIIPEVDLYKCEPWLPEKSFLLPGNDME
WYFYSTRDKKYPNGSRTNRATRAGYWKATGKDRTVESKMKMGKKTLVYYRGRAPHGLR
TNWVMHEYRLTHAPSSSLKESYALCRVFKNIQIPKRKGEBEABEESTSVGKEEBEKE
KKWRKCDGNYIEDESLKRASAESSSELTQGVLLDEANSSSIFALHFSSSLDDHDLFS
NYSHQLPYHPPLQLQDFPQLSMNEAIEIMSIIQQDFQCRDSMNGTLDEIFSSSATFPASL*

>G979 (60..1352)

CCTCTGAGGAATCAATCACTCACTCCAAAAAATCTAACTTTCTCAGAGTTTAA
TGAAGAAGCGCTTAACCACTTCCACTTGTCTTCTTCTCCATCTTCTCTGTTTCTTCTT
CTACTACTACTTCTCTCTCTATTAGTCCGAGGCTCCAAGGCCTAAACGAGCCAAAAGGG
CTAAGAAATCTTCTCTTCTGGTGATAAATCTCATAACCCGACAAGCCCTGCTTCTACCC
GACGCAGCTCTATCTACAGAGGAGTCACTAGACATAGATGGAGTGGGAGATTCGAGGCTC
ATCTTTGGGACAAAAGCTCTTGGAAATCGATTGAGAACAGAAGGCAACAAGTTTATC

TGGGAGCATATGACAGTGAAGAAGCAGCAGCACATACGTACGATCTGGCTGCTCTCAAGT
ACTGGGGACCCGACACCATCTTGAATTTTCCGGCAGAGACGTACACAAAGGAATTGGAAG
AAATGCAGAGAGTGAACAAAGGAAGAATATTTGGCTTCTCTCCGCCGCCAGAGCAGTGGTT
TCTCCAGAGGCGTCTCTAAATATCGCGGCGTCGCTAGGCATCACCACAACGGAAGATGGG
AGGCTCGGATCGGAAGAGTGTGGGAACAAGTACTTGTACCTCGGCACCTATAATACGC
AGGAGGAAGCTGCTGCAGCATATGACATGGCTGCGATTGAGTATCGAGGCGCAAACGCGG
TTACTAATTTTCGACATTAGTAATTACATTGACCGGTAAAGAAGAAAGGTGTTTTCCCGT
TCCCTGTGAACCAAGCTAACCATCAAGAGGGTATCTTGTGTAAGCCAAACAAGAAGTTG
AAACGAGAGAAGCGAAGGAAGAGCCTAGAGAAGAAGTGAACAACAGTACGTGGAAGAAC
CACCCGCAAGAAGAAGAAGAGAAGGAAGAAGAGAAAGCAGAGCAACAAGAAGCAGAGATTG
TAGGATATTCAGAAGAAGCAGCAGTGGTCAATTGCTGCATAGACTCTTCAACCATAATGG
AAATGGATCGTGTGTGGGACAAACAATGAGCTGGCTTGGAACTTCTGTATGATGGATACAG
GGTTTTCTCGTTTTTGGACTGATCAGAATCTCGGAATGAGAATCCCATAGAGTATCCGG
AGCTATTCAATGAGTTAGCATTTGAGGACAACATCGACTTCATGTTGATGATGGGAAGC
ACGAGTGCCTTGAACCTGGAAAATCTGGATTGTTGCGTGGTGGGAAGAGAGAGCCCACCCT
CTTCTTCTTCAACATTGTCTTGTCTTATCTACTGACTCTGCTTCATCAACAACAACAACAA
CAACCTCGGTTTTCTTGTAACTATTGGTCTGAGAGAGAGAGCTTTGCCTTCTAGTTTGAA
TTTTCTATTCTTCCGCTTCTTCTTCTTTTTTTCTTTTGTGGGTCTGCTTAGGGTTTG
TATTTCAAGTTTCAAGGGCTTGTTCGTGGTTCTGAATAATCAATGTCTTTGCCCTTTTTNN
AANGNTNCAAGNTNAAANAAAAAAAAA

>G979 Amino Acid Sequence (domain in AA coordinates: 63-139,165-233)

MKKRLTSTCSPSSSVSSSTTSSPIQSEAPRPKRAKRAKSSPSGDKSHNPTSPAST
RRSSIIYRGVTRHRWTGRFEHLWDKSSWNSIQNKKGKQVYLGAYDSEAAAHTYDLAALK
YWGPDITLNPFAETTYTKEEEMQRTVTKEEYLAQLRRQSSGFSRGVSKYRGVARHHNHRW
EARIGRVFGNKYLYLGTNTQEEAAAAYDMAAIEYRGANAVTNFIDISNYIDRLKKKGVPF
FPVNQANHQBEGILVEAKQEVETREAKEEPREEVKQQYVEPPQEEEEKEEKAEEQAEI
VGYSEEAUVNCCIDSSSTIMEMDRGDNNELAWNFCMMDTGFSPLTDQNLANENPIEYP
ELFNELAFEDNIDFMFDDGKHECLNLENLDCCVVGRESPPSSSSPLSCLSTDSASSTTTT
TTSVSCNYLV*

>G987 (1..4011)

ATGGGTTCTTACTCAGCTGGCTTCCCTGGATCCTTGGACTGGTTTGATTTTCCCGGTTTA
GGAAACGGATCCTATCTAAATGATCAACCTTTGTTAGATATTGGATCTGTTCTCCTCCT
CTAGACCCATATCCTCAACAGAATCTTGCTTCTGCGGATGCTGATTTCTCTGATTCTGTT
TTGAAGTACATAAGCCAAGTTCTTATGGAAGAGGACATGGAAGATAAGCCTTGTATGTTT
CATGATGCTTTATCTCTTCAAGCAGCTGAGAAGTCTCTCTATGAAGCTCTCGGCGAGAAG
TACCCGGTTGATGATTCTGATCAGCCTCTGACTACTACTAGCCTTGCTCAATTGGTT
AGTAGTCTGCTGGTCTTCTTCTTATGCTTCAAGCACCACAACCACTTCTCTGATTCACAA
TGGAGTTTGTATGTTTGGAGAATAATAGGCCTTCTTCTTGGTTGCAGACACCGATCCCG
AGTAACCTTCAATTTTTCAGTCTACATCTACTAGAGCCAGTAGCGGTAACGCGGTTTTCGGG
TCAAGTTTTAGCGGTGATTTGGTTTCTAATATGTTTAAAGATACTGACTTGGCGTTACAA
TTCAAGAAAGGGATGGAGGAAGCTAGTAAATTCCTTCTTAAGAGCTCTCAGTTGGTTATA
GATAACTCTGTTCTTAACAGATTAAACCGGAAAGAAGAGCCATTGGCGCGAAGAAGACAT
TTGACTGAAGAAAGAAGTAAGAAACAATCTGCTATTTATGTTGATGAACTGATGAGCTT
ACTGATATGTTTGACAAATTTCTGATATTTGGCGAGGCTAAGGAACAACCTGTATGCATT
CTTAACGAGAGTTTTCCCTAAGGAACCTGCGAAAGCTTCAACGTTTAGTAAGAGTCCTAAA
GGCGAAAAACCGGAAGCTAGTGGTAACAGTTATACAAAAGAGACACCTGATTTGAGGACA
ATGCTGGTTTTCTGTGCTCAAGCTGTTTCGATTAAACGATCGTAGAACTGCTGACGAGCTG
TTAAGTCGGATAAGGCAACATTCTTCATCTTACGGCGATGGAACAGAGAGATTGGCTCAT
TATTTTGCTAACAGTCTTGAAGCACGTTTGGCTGGGATAGGTACACAGGTTTATACTGCC
TTGTCTTCCAAGAAAACATCTACTTCTGACATGTTGAAAGCTTATCAGACATATATATCA
GTCTGTCCGTTCAAGAAAATCGCAATCATATTCGCCAACCATAGTATTATGCGGTTGGCT
TCAAGTGTCTAATGCCAAAACCATCCACATCATAGATTTTGGAAATATCTGATGGTTTCCAG
TGGCCTTCTCTGATTCTGACCTTGCTTGGAGACGTGGTTTCATCTTGTAAGCTTCGGATA
ACCGGTATAGAGTTGCCTCAACGTGGTTTTAGACCAGCCGAGGGAGTTATTGAGACTGGT
CGTCGCTTGGCTAAGTATTGTCAGAAGTTCAATATTCGTTTGGAGTACAATGCGATTGCG
CAGAAATGGGAATCAATCAAGTTGGAGGACTTGAAGCTAAAAGAAGGCGAGTTTGTGCG
GTAAACTCTTTATTTTCGGTTTAGGAATCTTCTAGATGAGACGGTGGCAGTGCATAGCCCG

AGAGATACGGTTTGAAGCTGATAAGGAAGATAAAGCCAGACGTGTTTCATCCCCGGGATC
CTCAGCGGATCCTACAACGCGCCTTTCTTTGTACAGAGGTTTAGAGAAGTTCTGTTTCAT
TACTCATCTCTGTTTGACATGTGTGACACGAATCTAACACGGGAAGATCCAATGAGGGTT
ATGTTTGAGAAAGAGTTCTATGGGCGGGAGATCATGAACGTGGTGGCGTGTGAGGGGACG
GAGAGAGTGGAGAGGCCAGAGAGTTATAAGCAGTGGCAGGCGAGGGCGATGAGAGCCGGG
TTTAGACAGATTCCGCTGGAGAAAGAACTAGTTCAGAACTGAAGTTGATGGTGGAAAGT
GGATACAAACCCAAAGAGTTTGATGTTGATCAAGATTGTCACTGGTTGCTTCAGGGCTGG
AAAGGTAGAATTGTATACGGTTTCATCTATTTGGGTTCTTTCTTTTCTATGTGGGCAGA
GCAACTAGGGTTTTGATCATGGATCCAACTTCTCTGAATCTCTAAACGGCTTTGAGTAT
TTTGATGGTAACCTAATTTGCTTACTGATCCAATGGAAGATCAGTATCCACCACCATCT
GATACTCTGTTGAAATACGTGAGTGAGATTCTTATGGAAGAGAGTAATGGAGATTATAAG
CAATCTATGTTCTATGATTCAATTGGCTTTACGAAAACTGAAGAAATGTTGCAGCAAGTC
ATTACTGATTCTCAAAATCAGTCTTTAGTCTGCTGATTCAATTGATTACTAATTCTTGG
GATGCAAGCGGAAGCATCGATGAATCGGCTTATTCGGCTGATCCGCAACCTGTGAATGAA
ATTATGGTTAAGAGTATGTTTGTAGTATGCAGAATCAGCTTTACAGTTAAGAAAGGGGTT
GAAGAAGCTAGTAAATTCCTTCCCAATAGTGATCAATGGGTTATCAATCTGGATATCGAG
AGATCCGAAAGGCGCGATTTCGGTTAAAGAAGAGATGGGATTGGATCAGTTGAGAGTTAAG
AAGAATCATGAAAGGGGATTTGAGGAAGTTAGGAGTAGTAAGCAATTTGCTAGTAATGTA
GAAGATAGTAAGGTTACAGATATGTTTGATAAGGTTTGTCTTCTGACGGTGAATGCGAT
CCGCAACATTGTTAGACAGCGAGATTCAGCGGATTTCGGAGTAGTAAGAACATAGGAGAG
AAAGGGAAGAAGAAGAAGAAGAAGAGTCAAGTGGTTGATTTTTCGTACACTTCTCACT
CATTGTGCACAAGCCATTTCCACAGGAGATAAAACCACGGCTCTTGAGTTTCTGTTACAG
ATAAGGCAACAGTCTTCGCTCTCGGTGACGCGGGGCAAAGACTAGCTCATTGTTTCGCT
AACGCGCTTGAAGCTCGTCTACAGGGAAGTACCGGTCCTATGATCCAGACTTATTACAAT
GCTTTAACCTCGTCTGTTGAAGGATACTGCTGCGGATACAATTAGAGCGTATCGAGTTTAT
CTTTCTCGTCTCCGTTTGTACCTTGATGTATTTCTTCTCCATCTGGATGATTCTTGAT
GTGGCTAAAGATGCTCCTGTTCTTCATATAGTTGATTTGGGATTCTATACGGGTTTCAA
TGGCCGATGTTTATTCAGTCTATATCAGATCGAAAAGATGTACCGCGGAAGCTGCGGATT
ACTGGTATCGAGCTTCTCAGTGCGGGTTTCGGCCCGCGGAGCGAATAGAGGAGACAGGA
CGGAGATTGGCTGAGTATTGTAAACGGTTTAAATGTTCCGTTTGAGTACAAAGCCATTGCG
TCTCAGAACTGGGAAACAATCCGGATAGAAGATCTCGATATACGACCAAACGAAGTCTTA
GCGGTTAATGCTGGACTTAGACTCAAGAACCTTCAAGATGAAACAGGAAGCGAAGAGAAT
TGCCCCGAGAGATGCTGTCTTGAAGCTAATAAGAAACATGAACCCGGACGTTTTCATCCAC
GCGATTGTCAACGGTTCATTCAACGCACCCTTCTTTATCTCGCGGTTTAAAGAAGCGGTT
TACCATTACTCCGCTCTCTTCGACATGTTTGATTGACGTTGCTTCGGGATAACAAAGAG
AGGATTAGGTTTCGAGAGGGAGTTTACGGGAGAGAGGCTATGAACGTGATAGCGTGCGAG
GAAGCTGATCGAGTGGAGAGGCTGAGACTTACAGGCAATGGCAGGTTAGAATGGTTAGA
GCCGGGTTTAAAGCAGAAACGATTAAGCCTGAGCTGGTAGAGTTGTTTAGAGGAAAGCTG
AAGAAATGGCGTTACCATAAAGACTTTGTGGTTGATGAAAATAGTAAATGGTTGTTACAA
GGCTGGAAGGTCGAACCTCTCTATGCTTCTTCTGTTGGGTTCTTGCCTAG

>G987 Amino Acid Sequence (domain in AA coordinates: 428-432,704-708)
MGSYSAGFPGLDWFDFPGLNGSYLNDQPLLDIGSVPPPLDPYPQQNLASADADFSDSV
LKYISQVLMEEDMEDKPCMFHDALSLQAAEKSLYEALGEKYPVDDSDQPLTTTSLAQLV
SSPGSSSYASSTTTTSSDSQWSFDCLNRPSSWLQTPIPSNFIFQSTSTRASSGNAVFG
SSFSGDLVSNMFNDTDLALQFKKGMEASKFLPKSSQLVIDNSVPNRLTGKKSHWREEEH
LTEERSKKQSAIYVDETDELTFMFDNLIIFGEAKEQPVCILNESFPKEPAKASTFSKSPK
GEKPEASGNSYTKETPDLRLTMLVSCAQAVSINDRRTADELLSRIRQHSSSYGDGTERLAH
YFANSLEARLAGIGTQVYALSSKKTSTSDMLKAYQTYISVCPFKKIAIIFANHSIMRLA
SSANAKTIHIIDFGISDGFQWPSLIHRLAWRRGSSCKLRITGIELPQRGFRPAEGVIETG
RRLAKYCQKFNIPFEYNALIAQKWESIKLEDLKLKEGEFVAVNSLFRFRNLLDETVAVHSP
RDTVLKLRIRKIPDVFIPGILSGSYNAPFFVTRFREVLPHYSSLFDMCDTNLTREDPMRV
MFEKEFYGREIMNVVACEGTERVERPESYKQWQARAMRAGFRQIPLEKELVQKLKLMVES
GYKPKFEFDVDQCHWLLQGWKGRIVYGSSIWVPPFFVYVGRATRVLIMDPNFPSESLNGFEY
FDGNPNLLTDPMEDQYPPPSDTLLKYVSEILMEESNGDYQSMFYDSLALRKTEMLQQV
ITDSQNQSFSPADSLITNSWDASGSIDESAYSADPQPVNEIMVKSMFSDAESALQFKKGV
EEASKFLPNSDQWVINLDIERSERRDSVKEEMGLDQLRVKKNHERDFEEVRSSKQFASNV
EDSKVTFMDFKVLILLDGECDPQTLLDSEIQAIRSSKNIGEGKGGKKKKKSQVVDFTLLT

HCAQAISTGDKTTALEFLLQIRQSSPLGDAGQRLAHCFANALEARLQGSTGPMIQTYYN
ALTSSLKDTAADTIRAYRVYLSSSPFVTLMYFFSIWMILDVAKDAPVLHIVDFGILYGFQ
WPMFIQSIISDRKDVPRLRITGIELPQCGFRPAERIEETGRRLAEBYCKRFNVPFEYKAIA
SQNWETIRIEDLDIRPNEVLAVNAGLRRLKNLQDETGSEENCPRDAVLKLRNMNPDVFIH
AIVNGSFNAPFFISRFKEAVYHYSALFDMFDSTLPRDNKERIRFEREFYGREAMNVIACE
EADRVPERPETYRQWQVRMVRAGFKQKTIKPELVELFRGKLKKWRYHKDFVVDENSKWLLQ
GWKGRTLYASSCWVPA*

>G993 (6..1091)

CAAAATATGGAATACAGCTGTGTAGACGACAGTAGTACAACGTCAGAATCTCTCTCCATCT
CTACTACTCCAAAGCCGACAACGACGACGAGAGAAGAACTCTCTTCTCCGCCGGCGACGT
CGATGCGTCTCTACAGAATGGGAAGCGCGGAAGCAGCGTCGTTTTGGATTACAGAGAACG
GCGTCGAGACCGAGTCACGTAAGCTTCCTTCGTCGAAATATAAAGGCGTTGTGCCTCAGC
CTAACGGAAGATGGGGAGCTCAGATTTACGAGAAGCATCAGCGAGTTTGGCTCGGTACTT
TCAACGAGGAAGAAGAAGCTGCGTCTTCTTACGACATCGCCGTGAGGAGATTCGCGGCC
GCGACGCCGTCACCTAACTTCAAATCTCAAGTTGATGGAAACGACGCCGAATCGGCTTTTC
TTGACGCTCATTCTAAAGCTGAGATCGTGGATATGTTGAGGAAACACACTTACGCCGATG
AGTTTGAGCAGAGTAGACGGAAGTTTGTAAACGGCGACGGAACCGCTCTGGGTTGGAGA
CGGCGACGTACGGAACGACGCTGTTTTGAGAGCGCGTGAGGTTTGTTCGAGAAGACTG
TTACGCCGAGCGACGTGCGGAAGCTGAACCGTTTAGTGATACCGAAACAACACGCGGAGA
AGCATTTTCCGTTACCGGCGATGACGACGGCGATGGGGATGAATCCGTCTCCGACGAAAG
GCGTTTTGATTAACTTGAAGATAGAACAGGGAAGTGTGGCGGTTCCGTTACAGTTACT
GGAACAGCAGTCAAAGTTACGTGTTGACCAAGGGCTGGAGCCGGTTCGTTAAAGAGAAGA
ATCTTCGAGCCGGTGATGTGTTTTGTTTCGAGAGATCAACCGACGACCGGCAATTGT
ATATCCACTGGAAGTCCGGTCTAGTCCGGTTCAGACTGTGGTTAGGCTATTTCGAGTCA
ACATTTTCAATGTGAGTAACGAGAAACCAAACGACGTGCGAGTAGAGTGTGTTGGCAAGA
AGAGATCTCGGGAAGATGATTTGTTTTCGTTAGGGTGTTCGAAGAAGCAGGCGATTATCA
ACATCTTGTGACAAATCTTTTTTTTTGTTTTTTCTTCAATTTGTTTCTCTTTTTTCA
ATATTTTGTATTGAAATGACAAGTTGTAAATTAGGACAAGACAAGAAAAAATGACAACTA
GACAAAATAGTTTTTGTTTAAAAAATAAAAAA

>G993 Amino Acid Sequence (domain in AA coordinates: 69-134)
MEYSCVDDSSSTSESLSISTTPKPTTTTEKKLSSPPATSMRLYRMGSGGSSVVLDSNGV
ETESRKLPSKSKYKGVVPQPNRWGAQIYKQHVWLGTFFNEEEEAASSYDIAVRRFRGRD
AVTNFKSQVDGNDAESAFDLDAHSKAEIVDMLRKHTYADEFEQSRRKFVNGDGKRSGLTA
TYGNDAVLRAREVLFEKTVTPSDVGKLNRLVIPKQHAKEHFPLPAMTTAMGMNPSPTKGV
LINLEDRTGKVVFRFRYSYWNSSQSYVLTKGWSRFVKEKNLRAGDVVCFERSTGPDRLQYI
HWKVRSSPVQTVVRLFGVNI FNVSNKPNDAVECVGKKRSREDDLFLGCSKKQAIINI
L*

>G681 (1..804)

ATGGGGAGGACGACATGGTTCGACGTCGACGGGATGAAGAAAGGAGAGTGACGGCAGAG
GAAGACCAGAAGCTCGGCGTTACATCAACGAGCATGGCGTTTGTGATTGGCGTTCCCTC
CCCAAAAGAGCTGGTTTGAGAGATGTGGAAAGAGCTGCAGATTAAGGTGGCTTAACAT
CTAAAGCCTGGGATTAGAAGAGGCAAATTCATCCTCAAGAAGAAGAAGAAATCATCCAA
CTTCATGCTGTTCTCGGAAACAGGTGGGCAGCATGGCGAAGAAGATGCAGAATCGAACA
GACAATGATATCAAGAACCATTGGAACCTTGTCTCAAGAAAAGACTTTCGAGAAAAGGGA
ATCGACCCATGACCCACGAGCCATCATCAAACACCTCACCGTCAATACCACTAACGCA
GATTGTGGTAACTCTTCCACCCAGCAGTCCCCGTCGACGACGGAAGCTCTCTTCTCC
GGCTCGTCTCGTCTTCTTAACTCGCCGAGGTATCTCATCTAGACAACATAGTCTC
GATAGGATCAAGTACATCTTGTGCAATTCAATAATCGAAAGCAGTGATCAAGCAAAAGAG
GAAGAAGAAAAAGAAGAAGAAGAAGAAGAGATTCAATGATGGGTGAGAAGATTGAC
GGTAGTGAAGGAGAAGATATTGAGATTTGGGGCGAGGAGGAAGTTAGGCGTTAATGGAG
ATTGATGCAATGATATGTACGAGATGACTTCGTACGACGCTGTGATGTACGAGAGTAGT
CACATACTTGATCATCTCTTTTACTTAATATAGTGTGACTGTGTGAGTGCATGCATGTT
>G681 Amino Acid Sequence (domain in AA coordinates: 14-120)
MGRTTWFVDGMMKKGEWTAEDQKLGAYINEHGVCDWRSLPKRAGLQRCGKSCRLRLWLN
LKPGIIRRGKFTPEEEEEIIQLHAVLGNRWAAAMAKMQNRDNDIKNHNWSCCLKRLSRKG
IDPMTHEPIIKHLTVNTNADCGNSSTTSPSTTESSPSSGSSRLNLKLAAGISSRQHS
DRIKYILSNSIISSDQAKEEEEEEEEEERDSMMGQKIDGSEGEDIIQIWGEEVRRIME

IDAMDMYEMTSYDAVMYESSHILDHLF*

>G1482 (1..996)

ATGAAGATCAGGTGCGACGTCTGCGATAAAGAAGAAGCGTCCGTGTTTTGCACGGCCGAC
GAAGCATCTCTGCGGGCGGCTGCGACCACCAAGTCCACCACGCTAACAACTCGCCTCT
AAACATCTCCGTTTCTCTCTCCTTTATCCTTCTTCTTCCAACACCTCCTCTCCTCTGCG
GACATCTGTTCAGGATAAAAAAGCTCTGTTGTTCTGTCAACAAGATAGAGCTATTTTATGC
AAAGATTGCGATTTCATCGATCCACGCTGCGAACGAACACACAAAGAAACACGATAGGTTT
CTTCTTACAGGGGTTAAGCTCTCTGCAACATCGTCTGTTTACAAACCTACTTCGAAATCT
TCTTCTTCTTCTTCAAGCAACCAAGATTCTCTGTCCCTGGATCATCAATCTCTAATCCT
CCTCCTCTCAAGAAACCTCTCTCAGCTCCTCCTCAGAGCAACAAGATCCAACCTTTTCG
AAGATCAACGGCGGTGATGCGTCCGTGAATCAGTGGGGATCCACAAGCACGATTCTTGAG
TATTTGATGGATACGTTACCTGGTTGGCAGGTTGAGGATTCTCTCGATTCTCTCTTCTCT
ACTTATGGTTTTCTAAGAGTGGTGATGATGATGGAGTGTACCATATATGGAACCAGAA
GATGACAACAACACTAAGAGAAACAACAACAACAACAACAACAACAACAATACAGTG
TCACTTCCATCTAAGAATTTAGGGATTGGGTCCCTCAGATTCCACAACTCTTCTTCT
TCATACCCAAATCAATACTTTTCTCAAGACAACAACATACAGTTTGGGATGTACAACAAA
GAAACATCACCAGAAGTAGTGTCTTTTGTCTCAATACAAAACATGAAACAACAAGGACAG
AACAACAAGAGATGGTATGATGATGGTGGCTTCACTGTCCACAGATCACTCCTCCTCCT
CTTCTTCTAATAAAAAAGTTTAGATCTTTCTGGTAA

>G1482 Amino Acid Sequence (domain in aa coordinates: 5-63)

MKIRCDVCDKEEASVFCTADEASLCGGCDHQVHHANKLASKHLRFSLLYPSSSNTSSPLC
DICQDKKALLFCQDRAILCKDCDSSIHAANEHTKKHDRFLLTGVKLSATSSVYKPTS
SSSSSSNQDFSVPGSSISNPPPLKKPLSAPPQSNKIQPFSSKINGGDASVNQWGSTSTISE
YLMDTLPGWHVEDFLDSSLPTYGFSKSGDDDGVLPMPEPEDDNTKRNNNNNNNNNNTV
SLPSKNLGIWVPQIPQTLPSYPNQYFSQDNNIQFGMYNKETSPEVVSFAPIQNMKQGGQ
NNKRWYDDGGFTVPQITPPPLSSNKKFRSFW*

>G225 (157..441)

CTCTCTCTCTCACTCTTTTCTTTTCCGAGAACCCACAAAAAAGCTACTATTAATCC
TTCCCTCGTGAGGAAATCATTTCTTCTTCTTCTCGAGATTTATCTCTTTCTCTCTCT
CTTTCTCTGTGTGTTTCTGTCTTTCAGATTAGTTCGATGTTTCGTTTCAGACAAGGCGGAA
AAAATGGATAAACGACGACGAGACAGAGCAAAGCCAAGGCTTCTGTTCCGAAGAGGTG
AGTAGTATCGAATGGGAAGCTGTGAAGATGTCAGAAGAAGAAGATCTCATTTCTCGG
ATGTATAAACTCGTTGGCGACAGGTGGGAGTTGATCGCCGGAAGGATCCCGGGACGGACG
CCGGAGGAGATAGAGAGATATGGCTTATGAAACACGGCGTCTGTTTTCGCAACAGACGA
AGAGACTTTTTTAGGAAATGATTTTTTTGTTTGGATTAAAGAAAAATTTCTCTCCTT
AATTCACAAGACAAGAAAAAAGGAAATGTACCTGTCTTGAATTACTATTTTGAATGT
ATAATTATCTATATATATAAGAAGAAAAAATGCTTAGGAATTT

>G225 Amino Acid Sequence (domain in AA coordinates: 39-76)

MFRSDKAEKMDKRRRRQSKAKASCSEEVSSIEWEAVKMSEEEEDLISRMVYKLVGDRWELI
AGRIPGRTPPEIERIERYWLMKHGVVFANRRRDFFRK*

>G226 (10..348)

CCAGTAGTTATGGATAATACCAACCGTCTTCGTCTTCGTGCGGTCCCAGTCTTAGGCAA
ACTAAGTTCAGTCCGATCCCGATATGACTCTGAAGAAGTGAGTAGCATCGAATGGGAGTTT
ATCAGTATGACCGAACAAGAAGATCTCATCTCTCGAATGTACAGACTTGTCGGTAAT
AGGTGGGATTTAATAGCAGGAAGAGTCGTAGGAAGAAAGGCAAATGAGATTGAGAGATAC
TGGATTATGAGAACTCTGACTATTTTCTCACAAACGACGACGTCTTAATAATTCTCCC
TTTTTTTCTACTTCTCCTCTTAATCTCCAAGAAAATCTAAAATTGTAAAGAAATCAAAAT
AAAAGCTTTCAATCTATAAAGTAGAACAAATCTTGAATGTCTTCTCA

>G226 Amino Acid Sequence (domain in AA coordinates: 28-78)

MDNTNRLRLRRGPSLRQTKFTRSRDYDSEEVSSIEWEFISMTEQEEDLISRMVYKLVGNRWD
LIAGRVVGRKANEIERIERYWIMRNSDYFSHKRRRLNNSPFFSTSPNLQENLKL*

>G9 (81..1139)

GTGTTTCTTCTTCTGCTAAAAGGTTATAATTTTGTCTTCTTGGTTGGTGAGAATCTTC
AAGAACTGAAACAAGAAAAATGGATTCTAGTTGCATAGACGAGATAAGTTCTCCACTT
CAGAATCTTTCTCCGCCACCACCGCCAAGAAGCTCTCTCCTCCTCCCGCGGCGGCTTAC
GCCTCTACCGGATGGGAAGCGGCGGAGCAGCGTCTGTTGGATCCCGAGAACGGCTAG
AGACGGAGTCACGAAAGCTACCATCTTCAAAATACAAAGGTGTTGTTCTCAGCCTAACG

GAAGATGGGGAGCTCAGATCTACGAGAAGCACCAACGAGTATGGCTCGGGACTTTCAACG
AGCAAGAAGAAGCTGCTCGTTTCCTACGACATCGCAGCTTGTAGATTCCGTGGCCGCGACG
CCGTGCTCAACTTCAAGAAGCTTCTGGAAGACGGCGATTAGCTTTTCTTGAAGCTCACT
CAAAGGCCGAGATCGTTCGACATGTTGAGAAAAACACACTTACGCCGACGAGCTTGAACAGA
ACAATAAACCGGCAGTTGTTTCTCTCCGTCGACGCTAACGAAAAACGTAACGGATCGAGTA
CTACTCAAACGACAAAGTTTAAAGACGTGTGAAGTTCTTTTCGAGAAGGCTGTTACAC
CTAGCGACGTTGGGAAGCTAAACCGTCTCGTGATACCTAAACAACACGCCGAGAAACACT
TTCCGTTACCGTCACCGTCACCGGCAGTGAATAAGGAGTTTGTATCAACTTGAAGACG
TTAACGGTAAAGTGTGGAGGTTCCGTTACTCATACTGGAACAGTAGTCAAAGTTACGTGT
TGACCAAGGGATGGAGTCGATTCTGCAAGGAGAAGAATCTTCGAGCCGGTGATGTTGTTA
CTTTCGAGAGATCGACCGGACTAGAGCGGCAGTTATATATTGATTGGAAGTTCCGGTCTG
GTCCGAGAGAAAACCCGGTTCAAGTGGTGGTTCGGCTTTTCGGAGTTGATATCTTAAATG
TGACCACCGTGAAGCCAAACGACGTCGTGGCCGTTTGGCGTGGAAAGAGATCTCGAGATG
TTGATGATATGTTTGGCTTACGGTGTTCCTCAAGAAGCAGGCGATAATCAATGCTTTGTGAC
ATATTTCTTTTCCGATTTTATGCTTTTCGTTTAAATTTTCTTTTGTCAAGTTGTGT
AGGTTGTGATTCATGCTAGGTTGTATTTAGGAAAAGAGATAAGACC

>G9 Amino Acid Sequence (domain in AA coordinates: 62-127)

MDSSCIDEISSSTSESFSATTAKKLSPPPAALRLYRMGSGGSSVLDPENGLTESRKL
PSSKYKGVVPQPNRWGAQIYEKHQVWLGTFFNEQEEAARSYDIAACRFRGRDAVVNFKN
VLEDGDLAFLEAHSKAEIVDMLRKHTYADELEQNNKRQLFLSVDANGKRNGSSTQNDKV
LKTCEVLFEKAVTPSDVVGKLNRLVI PKQHA EKHFPLPSPSPAVTKGVLINFEDVNGKVWR
FRYSYWNSSQSYVLTKGWSRFVKEKNLRAGDVVTFERSTGLERQLYIDWKVRSRGPENPV
QVVVRLFGVDIFNVTTVKPNDDVAVCGGKRSRDVDDMFALRCSKKQAIINAL*

>G1040 (51..863)

CTTTGATCTCCACTATTTAAGTAGACAAGAATCATAAAGAAAATAGTGAGATGATGATGT
TAGAGTCAAGAAACAGTATGAGAGCTTCAAACCTCAGTCCCAGATCTGTCTCTTCAGATCA
GTCTTCCTAACTATCACGCCGGAACCTCTTCACGGCGGTGACCGGAGCTCCACAAGCA
GTGATTCTGGAAGCAGCCTCAGTGACCTGAGCCATGAGAACAACCTCTTCAACAAACCTC
TCTTGAGCTTAGGATTTGACCATCATCATCAAAGGCGCTCAAACATGTTCCAACCTCAAA
TCTACGGTTCGAGATTTCAAGAGAAGCTCATCATCAATGGTTGGTCTTAAACGAAGCATT
GTGCTCCAAGAATGAGATGGACTTCTACTCTTCATGCTCACTTCGTCCATGCTGTTCAAC
TTCTTGGCGGCCATGAAAGAGCAACGCCCTAAATCAGTGTGGAGCTCATGAATGTGAAGG
ATCTAACCTTAGCTCATGTCAAGAGTCACTTGCAGATGTATAGAACAGTGAAATGCACTG
ATAAAGGATCACCAGGAGAAGGAAAGGTAGAGAAAGAGGCAGAGCAGAGGATAGAGGACA
ATAATAATAATGAAGAAGCTGATGAAGGAATGACACAAATTCGCCAACTCATCATCTG
TGCAAAAGACCCAAAGAGCTTCATGGTTCATCGACAAAGGAAGTATCTAGGAGCATATCTA
CAACAGCATATTTCACTTGGGAACAACCTCATCACTAAGGCCAATGAAGAGAAAGAGG
ATACCAACATTTCAATTTGGATTTACATTGGGCGGCCTAGTTGGGGGATGGAATA
TGCGGAACCTCCAGTGATTTAACCTTCTCAAGTGCTAATTGCCTTAAGCTACAACAAA
TAAGTCAGCTTAGGTTACAGTTTTAACATAATTTAACTTGTTTGATCATATGAGCTT
CGGAAGAATCATATTATCATCATATATGAACCTCTTTCCAAGAAATGTTCTATGAGTTTT
TGATATGTATAATCAAGAGAATCGTTTGAAGTAAAAA

>G1040 Amino Acid Sequence (domain in AA coordinates: 109-158)

MMLESRNMSRASNSVPDLSLQISLPNYHAGKPLHGGDRSSTSSDSGSSLSDLSHENNF
NKPLLSLGFDDHHQRSSNMFPQPIYGRDFKRSSSSMVGLKRSIRAPRMRWTSTLHAHFVH
AVQLLGGERATPKSVLELMNVKDLTLAHVKSHLQMYRTVKCTDKGSPGEGKVEKEAEQR
IEDNNNNEEADGTDTPNSPSSSVQKTQASWSSTKEVSRISITQAYSHLGTTHHTKANE
EKEDTNIHLNLDFTLGGVLVGGWNMRNPPVI*

>G2114 (64..1311)

ATAAAACGAAACCTATACATATAAACTAAGAGCGAGAAAGACAGCTAGAGAGAGAGAGA
GAGATGAAGAAATGGTTGGGATTTTCATTGACACCTCCTTTGAGAATCTGCAATAGTGAA
GAAGAAGAACTTAGGCATGACGGTTCCGATGTTTGGAGATATGATATTAACCTTTGATCAT
CATCATCATATGAAGACGTTCCAAAGGTGGAAGATCTCCTCTCAAACCTCATCAAACC
GAGTATCCTATAAAACCATAAACCAACCAATGTCAACTGCACCACTGTGGTTAACAGGTTA
AACCCACCCGGTTACCTTCTCCACGACCAACCCGTAGTTACACCACATTACCCGAACCTA
GATCCGAACCTTAGCAATGATTATGGAGGTTTTGAGAGGGTCGGTTCGGTCTCGGTTTTTC
AAATCTTGGTTAGAGCAAGGCACTCCAGCATTCCTCACTCTCGAGTCATTACGTTACTGAA

GAGGCTGGTACGAGCAATAATATTAGTCATTTTAGTAACGAAGAGACTGGTTATAACACC
AATGGCTCAATGCTATCATTGGCTTTGAGCCATGGGGCTTGTCTGATTTGATCAACGAA
TCGAATGTATCCGCACGGGTCGAAGAACCGGTTAAGGTAGATGAGAAGCGGAAGAGATTG
GTTGTTAAACCTCAGGTAAAGGAATCCGTTCCCTCGGAAGTCGGTTGATAGTTATGGACAA
AGAACTTCTCAGTATCGTGGAGTTACAAGGCATAGATGGACAGGGAGATATGAAGCTCAC
TTATGGGATAATAGCTGTAAGAAGGAGGGACAGACAAGGAGAGGAAGACAAGTGATCTT
GGAGGGTATGATGAGGAGGAGAAAGCAGCGAGGGCATATGATTAGCGGCTCTGAAGTAT
TGGGGTCTTACCCTCACTTAAATTTCCCTTTGAGTAATTACGAAAAGGAGATCGAGGAA
CTCAATAACATGAATCGGCAAGAATTTGTTGCCATGTTGAGGAGGAATAGCAGCGGGTTT
TCGAGGGGAGCTTCCGTGTATAGAGGAGTTACAAGGCATCATCAACATGGAAGGTGGCAA
GCCAGAATTGGAAGAGTTGCTGGAAACAAGGACTTGTACCTTGGAAACATTTAGCACGCAA
GAAGAAGCAGCGGAGGCGTACGATATCGCGCAATTAAATTCAGAGGCCCTAAACGCTGTA
ACCAATTTTCGATATAAATAGATATGACGTGAAGAGGATATGTTCAAGCTCAACGATTGTT
GATAGCGACAGGCCAAACATTCTCCACCAGCTCTGGCGCCGCCACTAACCGACACCG
TAAACTCCTCGCCGAGAGACTATCCACGTACGGTTGGTTTGGAGGAAATAAGTTCGTC
CAGTCTGTTTAAATCATTATGTTTAAATAACATATATTCCTAAGTAATTGAGGCCGTC
TACATATATACAACCTTTTGTAGCAAATTAAGTTATCAGAATCCACTATATATTCTCT
>G2114 Amino Acid Sequence (conserved domain in AA coordinates: 221-297, 323-393)
MKKWLGFSLTPPLRICNSEEEELRHGSDVWRYDINFDHHHDEDPVKVEDLLSNHQTE
YPINHNQTNVNCITTVNRLNPPGYLLHDQTVVTPHYPNLDPNLSNDYGGFERVGSVSVFK
SWLEQGTAPFPLSSHVYTEEAGTSNNISHFSNEETGYNTNGSMLSLALSHGACSDLINE
NVSARVEEPVKVDEKRRKRLVVKPQVKESVPRKSVDSYGQRTSQYRGVTRHRWTGRYEAHL
WDNSCKKEGQTRRGQVYLGYYDEEKAARAYDLAALKYWGPTHLNFPLSNYEKEIEEL
NNMNRQEFVAMLRNSSGFSRGASVYRGVTRHHQHGRWQARIGRVAGNKDLYLGTFTSTQE
EAAEAYDIAAIKFRGLNAVTFNFDINRYDVKRICSSSTIVDSDAQHSPTSSGAGH*
>G450 (65..751)
GAGTTATCGAGAGAGAGAGAAAAATATTTCTGATTTAAGACATATATAGACAGCAAGAAG
AGATATGAACCTTAAGGAGACGGAGCTTTGTCTTGGCCTCCCGGAGGCACTGAAACCGT
TGAAAGTCCGCGCAAGTCGGGTGTTGGGAACAGAGAGGCTTCTCCGAGACCGTTGATCT
CAAACCTTAATCTTCAATCTAACAAACAAGGACATGTGGATCTCAACACTAATGGAGCTCC
CAAGGAGAAGACCTTCTTAAAGACCTTCTAAGCCTCCTGCTAAAGCACAAGTGGTGGG
TTGGCCACCGGTGAGGAATACCGGAAAAATGTTATGGCTAATCAGAGAGCGGCGAAGC
AGAGGAGGCAATGAGTAGTGGTGGAGGAACCGTCGCCTTTGTGAAGGTTTCCATGGATGG
AGCTCCTTATCTTCGGAAGGTTGACCTCAAGATGTACACCAGCTACAAGGATCTCTCTGA
TGCTTGGCCAAAATGTTCAAGCTCCTTTACCATGGGGAGTTATGGAGCACAAGGGATGAT
AGATTTTCATGAACGAGAGTAAAGTGTATGGATCTGTTGAACAGTTCTGAGTATGTTCCAAG
CTACGAGGACAAAGATGGTGACTGGATGCTCGTTGGTGTATGTCCTTGGCCGATGTTTGT
CGAGTCATGCAACGTTTGGCGCATAATGAAAGGATCCGAAGCAATTGGACTTGTCTCCAAG
AGCAATGGAGAAGTTCAAGAACAGATCATGAACAAAAAAGAGGACAATATGCATTG
ATTTTTTTTTTTTTTGGTATTGTTATGATCATGTGTTTAAATTTAAATATAGGAAGGATA
TAGGAAAATATAATTGTTTACAAAAAATAACTTTAAATATGTCTTTTTTTTTTTTGA
AATTAGTCTGTGTTTTTGTTCATCTCTTAATTAGTAGAAATCATTTTTTAATATGTAA
TTGTGATAGTAAATCTATAGAGTTTCGTA
>G450 Amino Acid Sequence (domain in AA coordinates: TBD)
MNLKETELCLGLPGGTETVESPAKSGVGNKRGFSETVDLKLNLQSNKQGHVDLNTNGAPK
BKTFCLKDPSKPPAKAQVVGWPPVRNRYRKNVMANQKSGEAEAMSSGGGTVAFFVKVSMDGA
PYLRKVDLKMYSYKDLSDALAKMFSSFTMGSYGAQGMIDFMNESKVMDDLNSSEYVPSY
EDKGDWMLVGDVPWPMFVESCKRLRIMKGSEAIGLAPRAMEKFKNRS*
>G584 (40..1809)
AAAAAGTCTTCTCTTTTATAACTACGTCAGAGAAGTGTATGTCTCCGACGAATGTTCAA
GTAACCGATTACCATCTCAACCAATCAAAAACGGATACAACAAATCTCTGGTCAACCGAC
GACGATGCATCGGTAATGGAAGCTTTCATCGGCGGCGGCTCCGATCATCTCTCTTTTT
CCTCCACTTCTCTCTCTCTCTCTCAAGTCAACGAAGATAATCTCCAGCAACGTCTC
CAAGCTTTAATCGAAGGAGCAACAGGAACTGGACTTACGCCGTGTTCTGGCAATCATCT
CACGGTTTCGCCGAGAGAAGACAACAACAACAACACAGTGTGTTAGGTTGGGGAGAT
GGTTATTACAAAGGAGAAGAAGAGAAGTCTAGAAAGAAGAAATCAAATCCAGCTAGTGCA
GCTGAACAAGAGCATCGTAAGAGAGTGATTAGAGAGCTCAACTCTTTAATCTCCGGTGGT

GTAGGAGGAGGAGATGAAGCTGGAGATGAAGAAGTTACAGATACTGAATGGTTCTTCTTA
GTTTCAATGACACAGAGCTTTGTCAAGGGTACTGGTTTACCTGGTCAAGCTTTCTCAAAT
TCAGACACGATTTGGTTATCTGGTTCTAATGCTTTAGCTGGATCAAGTTGTGAGAGAGCT
CGTCAAGGTGAGATTTATGGGTTACAAACAATGGTGTGTGTAGCGACAGAGAATGGTGTC
GTTGAGCTTGGTTTCGTCGGAGATTATTCATCAAAGTTCAGATCTTGTTGATAAAGTTGAC
ACCTTTTTCATTTTAAACAATGGTGGTGGTGAATTTGGTTCTTGGGCGTTTAAATTTGAAT
CCAGATCAAGGAGAGAATGATCCAGGTTTGTGGATTAGTGAACCTAATGGTGTGACTCT
GGTCTTGTAGCTGCTCCGGTGATGAATAATGGTGGAAATGACTCAACTTCTAATCTGAT
TCTCAACCAATTTCTAAGCTTTGTAATGGAAGCTCTGTTGAAAACCTAACCCTAAAGTT
CTGAAATCTTGTGAAATGGTGAATTTCAAGAATGGGATTGAGAATGGTCAAGAAGAAGAT
AGTAGTAATAAGAAGAGATCACCGGTTTCGAATAATGAAGAAGGGATGCTTTCTTTTACC
TCTGTTCTTCCATGTGACTCGAATCACTCTGATCTTGAAGCTTCAGTGGCTAAAGAAGCT
GAGAGTAACAGAGTTGTGGTTGAACCGGAGAAGAAACCGAGGAACGAGGGAGAAAACCG
GCGAATGGAAGAGAAGAGCCCTTTGAATCATGTAGAGGCAGAGACAGAGAAGAGAGAAG
TTGAATCAGAGATTCATTTCTTTAAGAGCTGTGGTTCCTAATGTGTCTAAGATGGATAAA
GCTTCTCTATTAGGAGATGCTATTTCTGATATCAGTGAGCTTAAGTCTAAGTTGCAAAAG
GCTGAATCTGATAAAGAAGAGTTGCAGAAGCAGATTGATGTGATGAATAAAGAAGCGGGA
AATGCCAAAAGTTCCGTTAAAGATCGAAAATGTTTGAATCAAGAATCGAGTGTGTGATA
GAGATGGAGGTTGATGTGAAGATTATTGGTTGGGATGCAATGATAAGGATTCAATGTAGT
AAGAGGAATCATCTGGTGCTAAGTTTCATGGAAGCACTTAAGGAGTTGGATTGGAAAGTG
AATCATGCGAGTTTATCGGTAGTGAATGATCTTATGATCCAACAAGCGACTGTGAAAATG
GGGAATCAGTTTTTTACGCAAGATCAACTCAAGCTTGCTCTAACGGAGAAAAGTTGGAGAA
TGTCCATGAATTGAAGTCAGCATCTTTAGGGCTAATACACCGGAGAATACTGCGAAAAGT
CGAAAACAACGATCATAGTATAAGCCGCGTAAAAAGTGTTAAACCTTTCACACAAGTTT
CTCTAGTGAATGTAGTTGTAACTCTATTGTGTAAGGGTAATTTTGTAGTACCCACTTGT
TGCTATTGAATGCTTGTAGAGAGGATTCTTAGTGTAGTATATGATTAGGTTGGGGTTTG
TTGTTTTCATGAGATAAATAAATGTGTTGATCAATGGTTAAGTCTTTGGTTTGTGTTGT
ATGTATGTAAATAAGGCTTTTGTAGAAATAAGACAAATGGGACTGAAGTTGGAGTTTAA
AA

>G584 Amino Acid Sequence (domain in AA coordinates: 401-494)

MSPTNVQVTDYHLNQSKTD'TNLWSTDDASVMEAFIGGSDHSSLFPPLPPPPLPQVNE
DNLQORLQALIEGANENWTYAVFWQSSHGFAGEDNNNNNTVLLGWGDGYKGEEEKSRKK
KSNPASAAEQEHRKRVIRELNSLISGGVGGDEAGDEEVDTEWFFLVSMTQSFVKGTGL
PGQAFSNSD'TIWLSGSNALAGSSCERARQGQIYGLQTMVCVATENGVVLELGSSEIIHQSS
DLVDKVD'TFFNFNNGGGEFGSWAFNLNPDQGENDPGLW'ISEPNGVDSGLVAAPVMNNGGN
DSTNSDSQPI SKLCNGSSVENPNPKVLKSCMVNFKNGIENGQEEDSSNKKRSPVSNNE
BGMLSFTSVLP'CDNSHSDLEASVAKEASNRVVVEPEKKPRKRGRKPANGREEPLNHVEA
ERQRREKLNQRFYSLRAVVPNVSKMDKASLLGDATSYISELKS LQKAESDKEELQKQID
VMNKEAGNAKSSVKDRKCLNQESSVLIEMEVDVKIIGWDAMIRIQCKSRNHPGAKFMEAL
KELDLEVNHASLSVNDLMIQQATVKMGNQFF'TQDQLKVALTEKVGEC*.

>G668 (1..1056)

ATGGGAAGACCACCTTGCTGTGAAAAGATTGGAGTGAAGAAAGGGCCATGGACACCAGAG
GAAGACATCATCTTGGTTTCTTACATCCAAGAACATGGTCCTGGAAACTGGAGATCTGTC
CCAACACACACAGGTTTAAAGATGTAGCAAGAGCTGCAGATTGAGATGGACTAATTATCTT
CGACCCGGTATTAAGCGTGGAAATTTTACTGAGCATGAAGAGAAGACAAT'TGTTTATCTT
CAAGCCCTTTTAGGCAACAGATGGGCAGCCATAGCATCATACCTTCCAGAAAGGACAGAC
AATGATATAAAGAACTATTGGAACACTCACTTGAAGAAGAAGCTCAAAAAGATTAATGAA
TCTGGTGAAGAAGATAATGATGGTGTCTCTTTCATCAAACACTAGTTTCAAAAAGAACCAT
CAAAGCACTAACAAAGGTCAATGGGAAAGAAGACTTCAGACAGACATTAACATGGCAAAA
CAAGCTCTTTGTGAGGCCTTGTCTTTAGACAAACCATCATCCACTCTTTTCATCATCTTCA
TCATTACCGACACCAGTAATCACACAACAAAACATCCGTAACCTTCTCATCAGCTTTGCTT
GACCGTTGTTATGATCCATCCTCTTCTTCTTCTATCTACCACAACCACCCTACAAGCAAC
ACTACTAATCCATACCCATCAGGGGTATATGCGTCAAGTGCTGAGAACATCGCCCGGTTG
CTTCAAGATTTTCATGAAAGACACACCCAAAGGCTTTAACTTTATCATCTTCTCCTCGGTT
TCAGAGACTGGACCACTCACTGCTGCAGTCTCGGAAGAAGGTGGAGAAGGGTTTGAACAA
TCTTTCTTCACTTCAATTCAATGGACGAAACTCAAACTTGAAGTCAAGGAGACAAGCTTC
TTCCATGATCAAGTGATCAAACCGGAAATAACAATGGACCAAGATCATGGTCTAATATCA

CAAGGGTCTCTGTCTTTGTTTGAGAAATGGTTATTTGATGAGCAAAGCCACGAGATGGTT
GGTATGGCACTAGCAGGACAAGAAGGGATGTTCTAG

>G668 Amino Acid Sequence (domain in AA coordinates: 13-113)
MGRPPCCEKIGVKKGPWTPEDIILVSYIQEHGPGNWRVPTHTGLRCSKSCRLRWNTNYL
RPGIKRGNFTEHEEKTIVHLQALLGNRWAAIASYLPERTDNDIKNYWNTHLKKLKKINE
SGEEDNDGVSSSNTSSQKNHQSTNKGQWERRLQTDINMAKQALCEALSLDKPSSSTLSSSS
SLPTPVITQQNIRNFSSALLDRCDPSSSSSSTTTTTNTNPNYPGVIYASSAENIARL
LQDFMKDTPKALTSSSSPVSETGPLTAAVSEEGEGFEQSFNFNSMDETQNLTOETSF
FHDQVIKPEITMDQDHGLISQGSLSLFEKWLFDQSHMVGMLAGQEGMF*

>G1050 (23..1582)

TTCCCCATTTCAGAAAATCAAAATGGGTGGTGGTGGTGATACAACAGATACCAATATGAT
GCAGAGAGTTAATTCTTCTTCTGGTACATCGTCTTCTTCGATCCCTAAACACAATCTTCA
CTTGAATCCTGCTCTTATCCGCTCTCACCATCACTTCGTCACCCTTTCACCGGAGCTCC
TCCACCGCGGATTCACCCATTCTCCTTACTCTCAGATCCCGCGACTTTACAACCTAG
ACATTCTCGCTCTATGTCGCAACCGTCTTCTTCTCTCTTGTATTGATTCATGCGCGCGTT
AAATCCTTCTGCTCCGTCGGTTTCGGTGTGGTGGAGGAGAAAACCGTGCCGGATTTAG
TCCTTCGTTGCCTCCGTCACCGTTTACGATGTGTCTTCTTCTAGCTCTAGGAACGCCGG
AGATGGAGAGAATCTACCTCCGAGAAAGTCGCATAGGCGTTCGAATAGTGATGTTACTTT
TGGGTTTAGTTCAATGATGTCTCAGAATCAAAAGTCTCTCTTGTAGTTCTTTGGAGAG
ATCGATCTCTGGTGAAGATACATCAGATTGGTCTAATTGGTGAAGAAAGAACCGAGAGA
AGGCTTCTACAAGGGAAGAAAACAGAGGTTGAAGCAGCTATGGACGATGTTTTACGGC
TTATATGAATCTTGATAACATTGATGTCTTGAATCTTTTGGAGGTGAAGATGGCAAGAA
TGGGAATGAGAATGTGGAGGAGATGGAGAGTAGTAGAGGTAGTGGTACAAAGAAGACGAA
TGGTGAAGTAGTAGTATTCTGAAGGAGATAGCAGTCCGAGTGGGAATGTGAAGGTTGC
GTTGAGTTCTTCTTCTTTCAGGCGTGAAGAGAAGAGCAGGTGGAGATATTGCTCTACTGG
TAGACATTACAGGAGTGTCTTATGGACAGTTGTTTCATGGGGAAGTTGAATTTTCGGCGA
CGAATCATCGCTAAAGCTTCCGCCCTTCTTCATCAGCTAAAGTTTCCCCAACCAATTCAGG
TGAAGGGAATTCAGTGCTTATAGTGTGAATTTGGAACAGTGAGTTTACTGTCAGCTGA
AATGAAGAAGATTGCAGCTGATGAGAACTCGCTGAGATTGTAATGGCTGACCCTAAGCG
TGTTAAAGAATCTTGGCGAACCCTGCTATCTGCTGCAGCTTCAAAGGAGCGGAAGACGCG
ATACATGGCAGAGTTGGAACACAAGGTGCAGACACTTCAGACTGAAGCTACTACATTATC
GGCTCAGCTCACACATTTGCAGAGAGATTCTATGGGTTGACAAACCAGAACAGTGAGCT
GAAGTTTCGCTTCTCAAGCTATGGAGCAGCAAGCACAACTCCCGGATGCTCTGTCTAGAGAA
ACTGAATGAAGAAGTCCAGCGGTTGAACTGGTGATAGGGGAGCCGAACCGCAGGCAAAG
TGGGAGCAGCAGCAGCAATCAAAGATGTCACTAAACCCGAGATGTTTCAGCAGCTTAG
CATAAGTCAGTTACAACACCAACAGATGCAGCATTCCAATCAGTGTAGCACAAATGAAGC
AAAGCACACTTCAAACGACTAGGGTAAGTAAACTGCGATCCGAGTTGTCTAGTTACAT
ATATGATAAGAATCTTTGTGTCAGAGTTCTGTTTTTGAAGTTTAAAGAAACATATATA
AAGATTATGTCCGGGAAATTTGATCATATTTCTGAAACATACACATATATATAGTGG
TAATGGAGGACTTTCTTCTGACCA

>G1050 Amino Acid Sequence (domain in AA coordinates: 372-425)

MGGGGDTTDTNMQRVNSSSGTSSSSIPKHNHLNLPALIRSHHFRHPFTGAPPPPIPIPI
SPYSQIPATLQPRHSRMSQPSSFFSFDSLPLNPSAPSVSVSVEEKTGAGFSPSLPPSP
FTMCHSSSSRNAGDGENLPPRKSHRRSNSDVTFGFSSMMSQNQKSPPLSSLERSISGEDT
SDWSNLVKKEPREGFYKGRKPEVEAAMDVFTAYMNLNDIDLNSFGGEDGKNGNENVEE
MESSRSGSGTKKTNGSSSDSEGDSSASGNVKVALSSSSSGVKRRAGGDIAPTGRHYRSVS
MDSCFMGKLNFGDESSLKLPSSSAKVSPINSGEENSSAYSVEFGNSEFTAAEMKKIAAD
EKLAIEIVMADPKRVKRI LANRVSAARSKERKTRYMAELEHKVQTLQTEATTLQAQLTHLQ
RDSMGLTNQNSSELKFRLLQAMEQQAQLRDALSEKLNNEVQRLKLVIGEPNRRQSGSSSES
KMSLNPMPFQQLSISQLQHQQMQHSNQCSTMKAKHTSND*

>G1463 (199..1209)

TATCCTTCGCAAGACCCTTCTCTATATAAGGAAGTTCATTTCAATTTGGAGAGGACACGC
TGACAAGCTGACTCTAGCAGATCTGGTACCGTCGACAGTTTGAGATTTGCTTCATCCGGT
TTTTTTATTTCTGCAAAATATGTCACTCTCTCCCATTTTGTTCATATATAATATGTTTG
AAGTTTGATCAACTTAGTATGCGTTTCTTTTTCTCTAGTTCCTCTGTTTCTTGGTCTGA
TTTAGTTTCGTTATGGCGGACACACTGCTCAACGCAGAAGACGAAGTAATAATCTCACGT
TATCTGAAGCCTATGATCGTTAACAGAGTATCATGGCTGATCTCTTCATCGAAGACGCA

GACGTGTTCAACAAGGATCCATATGTGAAGTTCCATGCTGAGATCCCTAGCTTCGTGATC
GTTAAACCACGAACAAAGGCTTGTGGTAAAACCGATGGATGTGATTCTGGGTTGCTGGAGG
ATCATTTGGTCGTGATAAGCTGATAAAGTCGGAGGAGACTGGTAAGATTCTAGGGTTCAAG
AAGATACTCAAGTTCTGCCATAAGTGGAAACCTAGAGAATACAAGAGAAGTTTGGTAAATG
GAAGAGTATAGGCTTACCAATAACTTCAACTGGAAGCAAGATCATGTGATTTGCAAGATT
CGGCTTTTGTGTTGAAGCAGAAATTAGTTTCTTGCTAGCCAAGCATTTCTACACTACATCA
GACTCACTTCCTCGAAATGTGCTGTTGCCAGCTTATGGATTCTGTTCAACGATAAAACAA
GAGGAGGACGAATTTTATCCGGTGACGATAATGATTTTCAAGGAAAAGATTGGCCTAGC
TACGTTACCAACAACGTGATTGTCTGCATCCATCGGAGCTTGTGAATGTTACGATGGG
AAGTTTCATGATAACCGAATCTGCATCTTCGCTAACAGGACTTGTGGTGTAAACCGATAAA
TGCAATGAAGGTTACTGGAAGATTAAGCACCGTGAGAAGCTGATCATGTACGGTACGGG
CAGACCATTGGTTGGAAAGAAAGTTTTTTCAGTTTTATGAAACGGAGAAAAGAAAGACATTTT
GGTAATGGAGAAGAAGTGAAGGTAACCTGGACTCTAAAAGAGTATAGGCTTACCAGAAAA
ATGAACAAGAATAAAGTGGTGTGCGTTATCAAGTATAAGGTAAAGTGTTCACGAGGATA
ACTAGCTAGGGACTTCTACTCTTGTTTTCATGATCGATGCGACCGCTCTAGACAGGCCTC
GTACCGGATCCTCTAGCTAGAGCTTTTCGTTCTGATCATCGGTTTCGACAACGTTTCGTCA
>G1463 Amino Acid Sequence (conserved domain in AA coordinates:9-156)
MRFFSLVPLFLGRFSFVMADTLLNAEDEVIISRYLKPMIVNRVSWPDLFIEDADVFNKD
PYVKFHABIPSFVIVKPRTKACGKTDGCDSGCWRIIGRDKLIKSEBTGKILGFKKILKFC
LKWKPREYKRSVLVMEFYRLTNFNWQDHVICKIRLLFEAEISFLLAKHFYTTSDSLPRN
VLLPAYGFCSPDKQEEDEFYPVTIMISEGKDWPSYVTNNVYCLHPSSELVNVHDGKFHDNG
ICIFANRTCGVTDKCNEGYWKIKHREKLIMSRYGQTIGWKKVFQFYETEKERHFGNGBEV
KVTWTLKEYRLTRKMNKNKVVCVIKYKVKCLPRITS*
>G1944 (236..1306)
TCGACCTTCCTAATTTCCAACCTCTGTTCTTAGCAATATATTTTTTCTCCAAAAATAATT
CTCAGTTTGATTTTCTTCTCTAGCTCTTAAGTATATTTCTTTGTGTTATTTATCTTTT
AATCCTTTAATCTCATCTTTGTTTATCTTTAATCAAAACCCAAAATTTACATGGGTTCTT
GAAAATCTAGAAGAAATAAAGGAAACATAACAAAAATAGAAAGAAAAAGAAGCTAATGGT
CTTAAATATGGAGTCTACCGGAGAAGCTGTTAGATCAACCACCGGTAACGACGGTGGTAT
TACGGTGGTTAGATCCGACGCGCGTCAAGTTTCCACGTAGCTCAAAGATCAGAAAGCTC
AAACCAATCTCCACCTCTGTCACTCCTCCTCCACCACAGCCATCGTCTCATCACACAGC
TCTTCCGCGCTGCAAATTTGACGCGGTGACGACTACGACTACGACGCGCCGCGATGGAAGG
TATCTCCGGTGGACTGATGAAGAAGAAGCGTGGACGGCCAAGGAAGTATGGACCGGACGG
GACTGTTGTAGCGTTATCTCTAAACCGATTTTCATCAGCGCGCGCGCGTTCGCATCTTCC
GCCGCCGAGTTTACACGTCATCGATTTCTCCGCTTCTGAGAAACGTAGCAAAGTGAAACC
AACGAACTCGTTTAAACAGAAACAAAGTATCATCACCAAGTTGAGAATTTGGGTGAATGGGC
TCCTTGCTCCGTCGGTGGTAAATTTACACCTCATATAATCACAGTCAACACCGGCGAGGA
TGTAACAATGAAGATAATCTCGTTTTCGCAACAAGGACCTCGCTCTATTTGTGTTCTGTC
AGCAAACGGTGTATTTTCAAGCGTTACACTTCGTGAGCCAGATTCCTCTGGCGGCACATT
GACATACGAAGGTCGGTTTGAGATATTATCATTATCCGGGTCAATCATGCCTAATGATTC
AGGCGGAACACGAAGTAGAACGGGAGGAATGAGTGTATCGTTAGCAAGTCCCGATGGACG
TGTAGTAGGCGGTGGCTCGCCGGTTTACTAGTAGCCGCGAGTCCGGTTCAGGTGGTTGT
AGGAAGTTTTTTAGCGGGCACTGACCATCAAGATCAGAAACCGAAAAGAACAACATGA
TTTCATGTTGTGCGAGTCTTACCGCTGCAATTCCTATCTCTAGTGCAGCTGATACCGGAC
AATCCATTCCGCTCTCGTCTCTTCCGGTCAATAATAATACATGGCAGACTTCTTTAGCTTC
CGATCCAAGAAACAAGCATACCGATATTAATGTCAATGTAACCTGAAATCCAATCTTTCT
CTGTATTTTCTGTTAACAAGTTTGATTTGGTTGTTTATCTACATTAGGATTTTACTAAAA
TGGTAGTATATTTATAGGGTTTTAGGGTCTTTATTTTGGTTCCACTGTTGTCACTTGTA
GGATA
>G1944 Amino Acid Sequence (domain in AA coordinates:87-100)
MVLNMESTGEAVRSTTGNDDGGITVVRSDAPSDFHVAQRSESSNQSPSTVTPPPPQSSH
TAPPLQISTVTTTTTAAAMEGISGGLMKKRGPRKYGPDGTVVVALSPKPISSAPAPSH
LPPSSHVIDFSASEKRSKVKPTNSFNRTKYHHQVENLGEWAPCSVGGNFTPHIITVNTG
EDVTMKIISFSQOGPRISCVLSANGVISSVTLRQPDSSGGTLTYEGRFEILSLSGSFPN
DSGGTRSRRTGMSVSLASPDGRVVGGLAGLLVAASPVQVVVGSFLAGTDHQDQKPKKNK
HDFMLSSPTAAPISSAADHRTIHSVSSLPVNNNTWQTSLASDPRNKHTDINVNT*
>G2383 (37..990)

GACCTCTTTGATCCCTTCATTCCCCATCAAACAACCATGTTTCCTTCTTTTCACTACTCAC
ATTCAAAGCCCCTAATTCTCACCATCACTACTCTTCGCCTTCTTTTCTTTCTCTTCCGAT
TTTCTTGAGAGTTTTGATGAATCCTTCTTGATAAACCAATTCTTGTACAGCAGCAAGAT
GTAGCAGCAAAATGTTGTGAATCTCCTTGGAATTTTGAAGAAGCTTGAGCTTAAGAAG
AAGAATGAGAAGTGTGTTGATGGAAGCACCTCACAAGAGGTTCAATGGAGAAGGACGGTC
AAAAAAAGGGACAGGCATAGTAAGATCTGCACGGCTCAAGGTCCTAGAGACCGGAGGATG
AGGCTGTCTCTTCAGATTGCTCGCAAGTTTTCTGATCTTCAAGACATGTTGGGTTTCGAC
AAGGCGAGCAAGACGATTGAATGGCTTTTTCTCAAATCAAAGACTTCCATCAAACAACCTT
AAAGAAAGAGTGGCTGCATCGGAAGGAGGAGAAAGGATGAACATCTCCAGGTTGATGAA
AAGGAAAAGGATGAGACACTGAAGTTGAGAGTCTCAAAGAGAAGAAACAAAGACTATGGAG
AGCTCTTTTAAGACTAAAGAGTCGAGAGAGAGAGCTAGAAAGCGAGCAAGAGAGAGAACA
ATGGCAAAGATGAAGATGAGATTATTTGAGACCTCGGAAACAATTTTCAATCCTCATCAA
GAACTAGAGAGATCAAGATAACCAATGGTGTACAATTACTAGAAAAGGAAAATAAAGAA
CAAGAATGGAGTAATACTAATGATGTTTACATGGTAGAGTATCAAATGGATTCTGTGAGC
ATCATAGAGAAGTTTCTTGAGCTAACCAGTGACTCTAGCTCCTCTTCCATTTTTGGTGAC
TCCGAGGAATGTTACACAAGTCTTAGTTTCAAGAGGTACAATTTTCAAGCAGCAGGTAAC
AGCAATGTGTTAACTAAAAACCTAATTGAGTAATGCAGTTTTGATTATATTAGCTTTT
TGGTAATTCCAGGAATGTCGACACCAAGGG

>G2383 Amino Acid Sequence (conserved domain in AA coordinates:89-149)

MFPSFITHIQSPNSHHYSSPSFPFSSDFLESFDESFLINQFLQDVAANVVESPWKF
CKKLELKKKNEKCVDGSTSQEVQWRRTVKKRDRHSKICTAQGPDRMRRLSLQIARKFFD
LQDMLGFDKASKTIEWLFSSKTSIKQLKERVAAASEGGGKDEHLQVDEKEKDETLKLRVS
KRRTKTMESSFKTKESRERARKRERERTMAKMKMRLFETSETISDPHQETREIKITNGVQ
LLEKENKEQWSNTNDVHMVEYQMSVSIIEKFLGLTSDSSSSSIFGDSEECYTSLSVR
GTISAAGNSNVLTKNPN*

>G571 (326..1708)

TAGCCGACCTCTCTTCTCTCTCTGAAAAAACACCAAAGGAGCTTTAAATGCTCCGTTA
CATAATCTCTATCTCTTTCCAAGATATAGAGAAAGGAAAATAATATACAAGAATTAAAA
GAAGGTATATCATCATCTCTCTAGCTAGTGATCAAAGCACCGTCATCATCATATATC
ATCAGCTTGCCTCAGAGGAGAGAACACATAAGAGAGATCGAAGATCAAAATCTATCTC
TCTTCATCATCTTCTGCTGTACTATCATATCACACGCTCTCTCAAACATCATCTATAT
ATAGACTTCTCTTCATCATCATCAAATGCAAGGTCATCACCAGAATCATCATCAACACTT
ATCATCATCTCCGCCACGCTTCCCATGGAACTTCATGAACAAAGATGGGTATGATAT
TGGAGAGATAGACCCATCACTCTTCTCTATCTTGATGGACAAGGACATCATGATCCTCC
ATCAACTGCTCCTTCTCTTTACATCATCATCACAACTCAGAAATTTGGCGATGAGACC
TCCAACATCGACGCTCAACATCTTCCATCTCAGCCTATGCACATAGAGCCACCTCCTTC
TCTACACACCAATACCGATAATACAAGATTAGTTCCGGCTGCTCAACCTAGTGGTTCCAC
TCGACCACTTCTGACCCGTCATGGACTTGACCAATCTTCTAGTTTCATCAACCTCC
TCAAGGTTCTAAATCCATCAAGAAGGAAGGGAACCGCAAGGGTCTTGCTCATCGGACCA
TGACATACCTAAATCGTCAGACCCTAAAACATTGAGAAGACTAGCACAAAACAGAGAAGC
AGCAAGAAAAAGCAGATTACGTAAGGCTTATGTTTCAAGCACTCGAGTCATGTAGGAT
CAAACCTGACCCAACTAGAACAAGAGATTCAACGGGCCAGATCCCAAGGCGTATTCTTTGG
AGGGTCTCTTATAGGAGGAGATCAACAGCAAGGTGGACTACCCATTGGCCCTGGCAACAT
CAGCTCTGAAGCAGCGGTGTTTCGATATGGAATATGCGAGGTGGCTGGAGGAGCAGCAGAG
GCTATTAAACGAACCTAAGGTTGGCAACACAAGAACACTTGTCGAGAACGAGCTTAGGAT
GTTTGTGGACACATGTTTAGCTCATTATGACCATTTGATTAACTCAAGGCTATGGTCGC
TAAGACCGATGTCTTCCACCTCATTTCTGGAGCATGGAAAACCTCAGCTGAACGTTGCTT
CTTGTGGATGGGTGTTTCCGTCCATCGGAGATCATTAAGGTGATTGTGAACCAGATAGA
ACCATTGACGGAGCAACAGATAGTTGGGATATGTGGGCTGCAACAGTCCACACAAGAGGC
CGAGGAGGCTCTCTCGCAAGGCTCGAGGCGTTGAATCAATCACTTTCCGATAGCATTGT
CTCTGACTCCCTCCCGCTGCCTCCGCACCACCTCCTCCTCATCTATCCAATTTTATGTC
ACACATGTCCTTAGCTCTCAACAAGCTCTCTGCTCTCGAGGGCTTCGTTCTCCAGGCGGA
TAATTTGAGGCACCAACGATCCATAGGCTGAACCAATTGTTGACGACCCGTCAGAAGC
ACGGTGTCTTCTAGCCGTTGCGGAGTACTTCCACCGTCTTCAAGCTCTAAGTTCTCTCTG
GCTAGCCCGTCTCGGCAAGATGGATAATACTAAAACAACTGATGAAGGAAACCAAAAC
AAAAACAAGAGAATAGGTTGATTAGTTAGCCGCCAGCTTGACCTCTTTATCATATATATC
GTCTCTCTACTCAAATACAGTGAATTAGGGAAAATTGTTTGGCTTCTTTTGGTATATG

ATTCTTACTATTATGTTTTTAATCAAGA

>G571 Amino Acid Sequence (domain in AA coordinates: 160-220)
MQGHHQNHQHLSSSSSATSSHGNFMNKDGYDIGBIDPSLFLYLDGQGHDDPPSTAPSPLH
HHHTTQNLAMRPPTSTLNIFFSQPMHIEPPSSSTHNTDNLRLVPAAQPSGSTRPASDPSM
DLTNHSQFHQPPQGSKSIKKEGNRKGLASSDHDIPKSSDPKTLRRLAQNREAARKSRLRK
KAYVQQLESCRIKLTQLEQEIQRARSQGVFFGGSLIGGDQQQGGPLIGPGNISSEAAVFD
MEYARWLEEQQRLNLNELRVATQEHLSSENELRMFVDTCLAHYDHLINLKAMVAKTDVFLHI
SGAWKTPAERCFLWMGGFRPSEIIKVIVNQIEPLTEQQIVGICGLQQSTQEAEALSQGL
EALNQSLSDSIVSDSLPPASAPLPPHLSNFMHMSLALNKL SALEGFVLQADNLRHQTIH
RLNQLLTTRQEARCLLAVA EYFHLQALSSSLWLARPRQDG*

>G636 (6..1814)

CGATGATGCAACTGGGTGGTGGTACTCCGACCCTACAGCGGCGGCTACAACCGTCACAA
CTGCTACAGCACCACCGCCACAATCAACAACAACGATTACAGCGGCAACAGAAGCAGCGG
CAGCAGCGGTGGGGCGTTTGGAGGTGTCGGAAGAGATGCACGACCGTGGGTGGAGGAA
ATCGTTGGCCCGCGCAGGAAACGCTAGCGTTGTTGAAAATACGATCTGACATGGGAATAG
CGTTTCGAGACGCTAGCGTTAAAGGTCCCTTATGGGAAGAGGTTTCTAGGAAAATGGCGG
AGCATGGTTACATAAGAAACGCAAGAAATGCAAAGAGAAATTCGAGAACGTTTACAAAT
ACCACAAACGAACCAAGAAGGTTCGTACCGGAAAATCCGAAGGCAAACTTATCGCTTCT
TTGATCAATTAGAAGCTCTCGAGTCTCAATCTACAACCTCACTCCACCATCATCAACAAC
AAACGCCCTCTTCGACCACGCAAAAACAACAACAACAACAACAACAACAACAACAGCT
CCATATTTTCAACTCCTCCTCCGTAACGACAGTTATGCCGACGCTTCTCTTCTCATCAA
TTCCTCCGTATACTCAGCAGATTAAATGTACCTTCGTTTCCAAACATCTCCGGTGATTTTC
TATCGGATAATTCTACATCGTCTTCGTCTTCTTATTCGACTTCTTCTGACATGGAGATGG
GTGGTGGAAGTGGACTACAAGGAAGAAAAGGAAGAGGAAATGGAAGGTGTTTTTCGAGC
GGTTGATGAAACAAGTAGTTGATAAACAGGAAGAGCTTCAACGCACATTCTTGGGAAGCTG
TTGAAAAGCGAGAACACAAGAGATTGGTTAGAGAAGAGTCTTGGAGAGTTCAAGAGATTG
CCAGAATCAACCGCGAGCAGGAGATCTTAGCTCAAGAACGCTCTATGTCCGCTGCAAAAG
ACGCTGCTGTTATGGCCTTTCTTCAAAAAGTGTGAGAGAAACAACCGAATCAGCCACAAC
CGCAGCTCAGCCGCAACAAGTTTCGACCATCAATGCAGCTTAATAACAACAATCAGCAGC
AACCGCCTCAACGGTCTCCTCCACCGCAACCTCTGCTCCGCTTCCGCAGCCAATTCAAG
CGGTTGTGTCGACGTTAGACACAACGAAAACGCACAATCGTGGTGATCAGAATATGACTC
CTGCAGCTTCAGCGAGCTCGTCGCGGTGGCCGAAAGTGAGATAGAAGCATTGATAAAGC
TGAGGACGAATCTTGATTGAAATATCAAGAAAACGGACCAAAAGGACCATGTGGGAAG
AGATATCAGCGGAATGAGAAGGTTAGGATTCAACAGGAAGTCAAAGAGATGCAAAGAGA
AATGGGAAAACATAACAAATACTTCAAGAAAGTCAAAGAGAGCAACAAGAAACGTCCCG
AAGATTCCAAGACTTGGCCTTACTTTACCAGCTTGATGCTTTATATAGAGAGAGGAACA
AATTCACAGCAACAACAACATTGCAGCTTCTTCTTCTCATCTTCCGGTCTTGTTAAACCGG
ATAATTCTGTTCCTTGATGGTCCAACCAGAGCAGCAATGGCCTCCGGCTGTAACGACTG
CGACAATACTCCCGCAGCGGCTCAGCCTGATCAGCAATCTCAGCCGTCGGAGCAGAACT
TTGATGATGAAGAAGGTACAGATGAAGAGTACGACGATGAAGATGAGGAAGAGGAGAATG
AAGAAGAGGAAGGAGGTGAGTTCGAGCTTGTGCCTAGCAATAACAACAACAACAAGACGA
CGAATAATCTGTAATGATGATGATTGAGTTCGAACCGGTTTGGTGGTGAAAGATTAGTA
ATCTTTTTTTAAGTTTTGATACAGAACATGAGAATTTAAATATTGGAGGGTTT

>G636 Amino Acid Sequence (domain in AA coordinates: 55-145, 405-498)

MQLGGGTPTTTAAATTVTTATAPPPQSNND SAATEAAAAAVGA FEVSEEMHDRGF GGNR
WPRQETLALLKIRSDMGIAFRDASVKGPLWEEVSRKMAEHGYIRNAKKCKEKFENVYKYH
KRTKEGRTGKSEGKTYRFFDQLEALESQSTTSLHHHQQTPLRPQNNNNNNNNNNSSSI
FSTPPPVTVMPTLPSSSIPPYTQQINVPSPFNISGDFLSDNSTSSSSYSTSSDMMEMGG
GTATTRKKRKRKWKVFFERLMKQVVDKQEELQRTFLEAVEKREHKRLVREESWRVQEIAR
INREHEILAQERSMSAAKDAVMAFLQKLSEKQPNQPQPQPQVPRPSMQLNNNNQQQP
PQRSPPPQPAPLPQPIQAVVSTLDTTKTHNRGDQNMTPAASASSSRWPKVEIEALIKLR
TNLDSKYQENGPKGPLWEBISAGMRRLGFNRSKRCKEKENINKYFKVKESNKKRPED
SKTCPIYFHLDALYRERNKFHSNNNIAASSSSGLVKPDNSVPLMVQPEQQWPPAVTTAT
TPPAAQPDQSQPSEQNFDDEEGTDEEYDDEDEEEENEEEGGEFELVPSNNNNNKTTN
NL*

>G878 (197..1738)

CAAAAAAATCTCTCCCATTAAGACTGCCCAAGAAATATTTTATACAAAATGAAAGA

GAGAAACACGACACGAATTTTGTATAATTAAGATTACACAAAAAAGTGTTAGAAAGAG
AAATATCTTCTTCTTTTTTCTGTGTGAGTTGGGTTTGTAAAGTTTATCCTTTTTGTTC
TCAAAATCAAGAATCGATGGCGGAGAAGGAAGAAAAAGAACATCGAAGTTAAATCATC
CACC GGAGTTTACGGCCAACGATTTCACTACCTCCTCGACCGTTTGGTGAAATGTTTTT
TAGCGGTGGCGTTGGATTTAGTCTTGACCAATGACTCTCGTCTCAAATTTATTCTCTGA
TCCTGATGAGTTTCAAGTCTTTCTCTCAGCTTTTAGCTGGAGCTATGGCTTCTCCGGCGGC
AGCTGCTGTTGCCGCCGCTGCTGTGGTTGCTACTGCTCATCATCAGACACCTGTGAGCTC
TGTCGGTGATGGCGGTGGAAGCGGTGGTGATGTTGACCCGAGGTTTAAGCAGAGTAGACC
AACGGGATTGATGATACTCAACCACCGGGGATGTTTACTGTACCGCCGGGGTTAAGTCC
GGCTACTCTTTTGGATTCTCCGAGCTTCTTTGGTCTTTTTTACCTCTTCAGGGAACATT
TGGTATGACACATCAACAAGCTTTAGCACAAAGTCACTGCACAAGCAGTTCAAGGCAATAA
TGTTTATATGACGAATCACAACAATCTGAATATCCTTCTTCTACACAACAACAACA
ACAACAACAAGCTTCAATGACTGAGATTCATCATTTTTCTTCTGCACCTAGGTCTCA
GATTCGAGCCTCGGTTCAAGAAACATCGCAGGGTCAGAGAGAGACTTCGGAATATCTGT
CTTTGAGCATCGGTACAGCCTCAAAATGCTGACAAAACAGCTGATGATGGATACAACTG
GCGGAATATGGGCAGAAGCAAGTGAAGGGGAGCGATTTTCTCGGAGTTATTACAAATG
TACGCATCCAGCTTGTCTGTCAAGAAGAAAGTGGAGAGGTCCTCGATGGACAAGTAAC
GGAAATCATCTACAAGGGTCAACACAATCATGAGCTTCTTCAAAGCGCGGTAAACAATAA
CGGGAGTTGTAAAGTTCTGATATTGCAAATCAGTTTCAAACAAGTAATAGCAGTCTCAA
CAAGAGTAAGAGGGACAGGAAACAAGCCAAGTTACAACAACAGAGCAGATGTCTGAAGC
AAGTGATAGCGAGGAGTTGGGAATGCAGAGACTAGTGTGGGAGAAAGACATGAGGATGA
GCCTGATCCCAAGCGAAGAAATACAGAAGTTCCGGTTTTCAGAACCAGTTGCTTCATCGCA
TAGAAGTGTGACAGAGCCTAGGATTATTGTCCAAACGACGAGTGAAGTTGACCTCTTAGA
TGATGGATATAGGTGGCGCAAGTATGGTCAGAAAGTAGTCAAAGGAAATCCTTATCCGAG
GAGCTACTATAAGTGTACAACACCAGATTGCGGAGTAAGGAAACATGTAGAGAGAGCAGC
AACTGACCCAAAAGCTGTTGTAAACAACATATGAAGGTAAACATAACCATGATGTTCCAGC
TGCTAGAACCAGCAGCCATCAGTTAAGACCAAACAATCAACACAACACCTCAACGGTTAA
CTTCAATCATCAACAGCCTGTTGCACGTTTAAAGCTTAAAGAAGAGCAAATCACTTGACA
GAGAAGAAGAATACGACGGCGCTTGAGCTTTTGTGAGTTTAAATGAATCTTCTTTTTGGTT
AATGAACCTGTTTTTGTGCTTCAAAACACCACAGGTTTCTCTGGACAGAATCTCTGATA
TTACAGTTTCAAAGGTATGTTCTTTTATTTTATGTTGGAATCTTCTGTGAATCTTAAG
AAGCTTTAGGAGGTAATGTAAAAAACAGATTCAAAGTTATGCCCTTATGTGAATCTTT
TGTACATGGGATAAAACAAATTTACAGGTATCCTTTTTGTTCTTGTGTAATAAAAAA
AAAA

>G878 Amino Acid Sequence (domain in AA coordinates:250-305, 415-475)

MAEKKEKEPSKLSSTGVSRPTISLPPRPFGEFFSGGVGFSPGPMTLVSNLFSDFDEFK
SFSQLLAGAMSPAAAATAAAVATAHHQTPVSSVGDGGSGGDVDPFKQSRPTGLMI
TQPPGMFTVPGLSPATLLDSFPLGLFSPPLQGTFGMT HQQALAQVTAQAVQNNVHMQQ
SQQSEYPSSTQQQQQQQQQASL TEIPSFSSAPRSQIRASVQETSQGGRETSEISVFEHRS
QPQNADKPADDGYNWRKYQKQVKGSDFPSYKCTHPACPVKKKVERSLDGQVTEIIYK
GQHNHELPQKRGNNGSCKSSDIANQFQTSNSSLNKS KRQETSQVTTTEQMSEASDSEE
VGNAETSVGERHEDEPDPKRRNTEVRVSEPVASSHRTVTEPRIIVQTTSEVDLLDDGYRW
RKYGQKVVKGNPYPRSYKCTTPDCGVKRKHVERAATDPKAVVTTYEGKHNHDVPAARTSS
HQLRPNNQHNTSTVNFNHQQPVARLRLKEEQIT*

>G1134 (61..849)

TAAAGAAAGAGAAAAAAGCTTTCGTAGTGTCTATTGAAACCAGAGAAAAGCCAAAGGGG
ATGCAACCAACATCCGTGGTAGTAGCGGCGGTGGTGACGACGAGGAGGCAGAGGAGGA
GGAGGAGGGCTAAGTAGAAGTGGACTATCTCGGATCCGTTTCACTCCAGCGACTTGGCTT
GAAGCTTTACTTGAGGAAGATGAAGAAGAGTCTTTGAAACCTAATCTTGGTCTCACCGAT
TTGCTTACCGGGAACCTCGAACGATTTACCGACAAGTCGCGGCTCGTTTCGAGTTCCCGATT
CCTGTTGAGCAAGGGTTGTATCAACAAGGTGGGTTTCAACGACAGAATAGTACTCCGGCG
GATTTTCTTAGTGGTTCTGATGGATTTATCCAAGCTTTGGGATTACAGCGAATTACGAT
TACTTATCGGGGAATATCGATGTTTCTCCGGGAAGTAAGCGGTCTAGAGAAATGGAAGCA
CTCTTCTCTTCTCTGATTTTACTTCTCAAATGAAAGGAGAGCAAAAGCAGCGGTCAAGTT
CCTACCGGAGTATCAAGCATGTCCGATATGAACATGGAGAACCTTATGGAGGACTCTGTT
GCTTTTAGGGTTCCGGCTAAACGTGGTTGCGCAACTCATCCCCGAGCATTGCCGAGAGG
GTACGAAGGACGCGGATTAGTGATCGGATAAGGAAGCTACAAGAGCTTGTACCTAACATG

GACAAGCAAACCAACACTGCAGACATGTTAGAAGAAGCAGTAGAATACGTGAAAGTTCTT
CAAAGGCAGATCCAGGAGTTAACAAGAAGAGAGGTGCACATGCATACCTAAGGAA
GAACAATAAGGTTTGCTCCTGATTTGTTTTATATTGCTTAACGGCAATGATCTGATCGA
AAAATTCGAAAGATGATCTTAGCTTGAATTTAGATGGATGTCATGTTGAAAAGTATATTA
TTTGATAAAATGGATGTAGGTGTAATAAATTTTTGTACAATAATGAAGAAAGTTAAAA
AGAATTAATGAAAACATATATTCTTTATGATATAAAAAAAAAA

>G1134 Amino Acid Sequence (domain in AA coordinates: 198-247)

MQPTSVGSSGGDDGGGRGGGGLSRGLSRIRAPATWLEALLEEDEEESLKPNLGLTD
LLTGNSNDLPTSRGSFEFPIPVEQGLYQQGGFHRQNSTPADFLSGSDGFIQSFGIQANYD
YLSGNIDVSPGSKRSREMEALFSSPEFTSQMKGEQSSGQVPTGVSSMSMDNMENLMEDSV
AFRVRAKRGKATHPRISIAERVRRTRISDRIRKLQELVPNMDKQTNATDMLEEAVEYVKVL
QRQIQELTEEQKRCTCIPKEEQ*

>G1008 (89..973)

GCCTTTTTGACTCTTCTTTCTCTCTTCTACTTTTTTTCAGGCTCTCTCTCTATATCTCTA
TCTTCTTCTCCGGTTAACTAAAAGAGAAATGAAAAGCCGAGTGAGAAAATCCAAGTACAC
GGTTCACCGGAAAATCACATCCACACCGTTTCGACGGTTTCCCGAAGATTGTCAAATCAT
AGTCACTGACCCATGCGCTACTGATTTCTCCAGCGATGAGGAAAACGACAACAAATCTGT
TGCTCCGAGGGTGAAAACGTTATGTGGATGAGATCAGGTTCTGTGACGAAGATGACGAACC
TAAACCGGCGAGGAAAGCGAAGAAAAGTCCCGGCGGCTGCGGCGGAGAACGGTGGAGA
TTTGGTAAAGTCTGTGGTGAAGTATAGAGGAGTGAGACAACGACCTTGGGGAAAATTTGC
GGCGGAGATTCTCGATCCTTCGAGTCGTACTAGACTCTGGCTTGGGACTTTTTCGACGGC
GGAGGAAGCTGCTATAGGTTACGATAGAGCCGCGATTTCGAATCAAAGGTCATAACGCTCA
GACGAATTTTCTCACTCCTCCTCTAGTCCGACGACTGAGGTGTTACCGGAAACTCCGGT
GATTGACCTTGAACTGTCTCTGGTTGTGATTTCGGCGAGGGAATCGCAAATCAGTCTGTG
TTCTCCGACTTCTGTTCTCCGGTTTAGTCACAACGACGAAACAGAGTACAGAACAGAGCC
AACGGAAGAACAAAATCCGTTTTTCTGCTGATTGTTTCGCTCCGGAGATTATTTTG
GGATTCCGAAATTACCCCTGACCCTTTGTCTCGACGAATTCACCAATCCTTGTACC
AAACATCAACAACAACAACAGTGTGTGATAAGGATACGAATCTGTCTGATAGTTTCC
GTTGGGAGTGATCGGAGATTTTCAGCTCATGGGATGTTGATGAGTTTTTCCAAGATCATT
GTTGGATAAGTAATTTGATGAGTTCTTCCCCAGAATTTTTCTGGGTTTTCTTTTTGGTT
GTGTGAGTGAGATGAGTGGTTTGTAGACAACGACGGGGATGAATCTTAGCCGTCGGTTTT
CCATTTCTGTGGACGGCTCCGATCAGCGGAAGAAGCGCAACGGAGTTTTTATTTATCTGTT
TGAGAATTTTATAATTTAATTTGCGAGTAAATATAGTAATTAGTGTTAAGATTGTGAGAG
TTTAAGTTAATTAGGGAGGGGTTTTGAATATTGGGGATTTGGGAGGTTTTTGTGTTGGTT
TCTCTCCAAGTCTGCTACTATGCAAGGAAGCAGTATAAAGACCGTATATATATTTATTA
TTAATATTGATAAAAGTAAAAAAAAAAAAAAAAA

>G1008 Amino Acid Sequence (domain in AA coordinates: 96-163)

MKSRVRKSKYTVHRKITSTPFDGPKIVKIIIVTDPATDSSSDEENDNKSVAAPRVKRYVD
EIRFCDEDEPKPARKAKKKSAAAAENGDLVKSVMKRGVRQRPWGKFAAEIRDPSSR
TRLWLGTAFATAEAAIGYDRAAIRIKGHNAQTNFLTTPPSPTTEVLPEPVIDLETVSGC
DSARESQISLCSPTSVLRFSHNDETEYRTEPTTEQNPFPLPDLFRSGDYFWDSEITPDPL
FLDEFHQSLLPNINMNTVCDKDTNLSDFSPLGVIGDFSSWDVDEFFQDHLDDK*

>G1020 (132..689)

CTGTTCAACAAGAAAGCTCCCCAAAAGGAGCGTTGCTTTACTCTCTCTATAAAAAGAAGCTC
TTCTACTTCTTCTCGTTACCACAAAACCTTTTACCGATCTTCTCGTTCCATTCTTCTTC
CTAATTACACCATGCCCAACATCACCATGGGTTTGAAACCGGACCCGGTTGCTCCAACGA
ACCCGACTCATCATGAGAGTAATGCTGCCAAAGAGATTCTGTTACAGAGGCGTTAGGAAAC
GTCCATGGGGAAGATACGCGCTGAGATCCGAGATCCGGTTAAGAAAACCTCGAGTCTGGC
TCGGTACGTTTCGACACCGCTCAGCAGGCGGCGCGTGCCTTACGACGCAGCCGCGCTGACT
TTCGTGGTGTAAAGGCTAAGACCAATTTCGGTGTATCTGTTGGTAGTAGTCTCTACTCAGA
GTAGCACCCTCGTCTGACTCTCCACGGCGGCACGGTTTATAACACCTCCGCACCTCGAGC
TCAGCTTAGGCGGCGGCGGCGCTGTCGTCGTAAGATCCCGCTTGTGCATCCGGTTTACT
ACTATAACATGGCGACGTATCCAAAGATGACGACGTGTGGTGTCCAGAGCGAGTCTGAAA
CGTCGTGGTCTGTTGATTTTGAAGGTGGAGCTGGGAAGATATCTCCGCCGTTAGATCTGG
ATCTTAACCTTAGCTCTCCGCGGAATAGGCCGTGAGTTTTTTTTTTCTTATGTCGTTTC
TTTAGACAAAAAAATAACGTTTCTTTTTTTTTCTGCCTAAGAAAAAATATTATCCG
TTTTTTAGAAGAAAAA

>G1020 Amino Acid Sequence (domain in AA coordinates:28-95)
MPNITMGLKPDFVAPTNPHTHESNAAKEIRYRGVRKRPWGRYAAEIRDVPVKTRVWLGT
DTAQQAARAYDAAARDFRGVKAKTNFGVIVGSSPTQSSTVVDSPTAARFITPPHLELSLG
GGGACRRKIPLVHPVYYYNMATYPKMTTCGVQSESETSSVVDFEGGAGKISPPLDLNL
APPAE*

>G1023 (252..1250)

TCGTCTTCTTAATCGCTTTCTGCTCTGTTTTCTCGTTCATCAAGCTACATCTACTAGCT
CTCTCAGTGATGATTTCTCACAGTTTCATCGATTTCCATGCGTTTAAAGACCTAAAAGGA
CTTGTTCTGGGGTAAAGGACTTTTCTTGTTCTTGAGAGAGTTCATTTTGAGGCTTTTCTG
GGAATTTTGAGAGGTTTTTATAGGGTTTAAAGGGGTTTGGTTTTGAATTTTCGCACACCAAG
TGTTTCGATAAAATGGCTGAACGAAAGAAACGCTCTTCTATTCAAACCAATAAACCAACA
AAAAACCCATGAAGAAGAAACCTTTTCAGCTAAATCACCTCCAGGTTTATCTGAAGATT
TGAAGACTATGAGAAAACCTCCGTTTCGTTGTGAATGATCCTTACGCTACTGACTACTCAT
CAAGCGAAGAAGAAGAAAGGAGTCAGAGAAGGAAACGTTATGTCTGTGAGATCGATCTTC
CTTTCGCTCAAGCTGCTACTCAAGCAGAATCTGAAAGCTCATATTGTCAGGAGAGTAACA
ATAATGGTGTAAAGCAAGACTAAAACTTCAGCTTGTAGCAAAAAGGTTTTACGCAGCAAAG
CATCTCCGGTCGTTGGACGTTCTTCTACTACTGTCTCGAAGCCTGTTGGTGTTAGGCAGA
GGAAATGGGGTAAATGGGCTGCTGAGATTAGACATCCAATCACCAAAGTAAGAACTTGGT
TGGGTACTTACGAGACGCTTGAACAAGCAGCTGATGCTTATGCTACCAAGAAGCTTGAGT
TTGATGCTCTGGCTGCAGCCACTTCTGCTGCTTCTCTGTTTGTCAAATGAGTCTGGTT
CTATGATCTCAGCCTCAGGGTCAAGCATTGATCTTGACAAGAAGCTAGTTGATTGCACTC
TTGATCAACAAGCTGGTGAATCGAAGAAAGCGAGTTTTGATTTGACTTTGCAGATCTAC
AGATTCTGAAATGGGTTGCTTCATTGATGACTCATTCAATCCCAAATGCTTGTGAGCTTG
ATTTTCTCTTAACAGAAGAGAACAAACCAAAATGTTGGATGATTACTGTGGCATAGATG
ATCTGGACATCATTTGGTCTTGAATGTGACGGTCCAAGCGAACTTCCAGACTATGATTTCT
CAGATGTGGAGATCGATCTTGGTCTCATTTGGAACCAACATTGACAAGTATGCTTTCGTTG
ATCATATCGCAACAACTACTCCCACTCCTCTTAATATCGCGTGCCCAATAAGTTTGCAGC
TAGGTGTTATTATTAGCTATAGGAGCAACGTAAAAAGCTCGTTGTTACTCGGTTTTGTCT
TAAGTTATTAAAGTATAGCAGAGGCAGTTAATCTCAAGGGAAGCAAAAACCTAAAGATA
GAAGCAGATGCAGTTTTGTGTGTTGGTGTACTAAAGAAAGTTTTGTTGACATAATGGTT
TTGATGTTGTGGAGAAGATAGAGAGGTGTGATCGAAATGTAAATCTCAGGTGGTTTTTT
TTGAAGGCAATTGTTTCTCATTTAGGGTTTTTTTCTATATGAGGATTGTCTTTGAAAAGC
CTTTAGATGTTTTCTAATTCGTAAGCTCTCTCAATCTTTGTAAGTTTTGCCTGTTGAGTT
ATTGATACATATGTGAGACCTACTTTATTTGTTTTGTGCTACATACATTGTTGATGGTTT
CGTCAAAAAAAA

>G1023 Amino Acid Sequence (conserved domain in AA coordinates:128-195)

MAERKKRSSIQTNKPNKKPMKKPFQNLHLPGLSEDLKTMRLRFVNDPYATDYSSSEE
EERSQRRKRYVCEIDLPPAQAATQAESESSYQESNNNGVSKTKISACSKKVLRSKASPV
VGRSSTTVSKPVGVRQRKWKWAAEIRHPITKVRTWLGTYETLEQAADAYATKKLEFDAL
AAATSAASSVLSNESGSMISASGSSIDLKKLVDSTLDQAGESKKASFDFADLQIPE
MGCFIDDSFIPNACELDLFLTEENNNQMLDDYCGIDDLDIIGLECDGPSELDPDYDFS
DVEIDLGLIGTTIDKYAFVDHIATTTPTPLNIACP*

>G1053 (38..538)

GAAACTCTTACATACTCATATAAAACCAAACCTAAACCATGATTCCGGCAGAAATCAACGG
ATATTTCCAATATCTATACCGGAATACAACGTAATAAACATGCCTTCATCTCCAACCTC
TTCCTTAAACTACCTAAACGATTTGATCATCAACAACAACAACTATTCTCATCATCCAA
CAGTCAAGATCTCATGATAAGCAACAACCTCAACTTCCGACGAAGATCATCATCAAAGCAT
CATGGTACTCGACGAGAGGAAACAGAGAAGGATGCTTTCGAACAGAGAATCTGCAAGGAG
GTCAAGGATGAGGAAACAGAGACATCTTGATGAACTCTGGTCTCAGGTAATAAGGCTTCG
CAACGAGAACAACCTGTCTTATCGATAAGCTGAACCGCGTATCGGAGACTCAAAATGTGT
ATTGAAGGAGAACTCTAAACTCAAAGAAGAAGCTTCTGATCTCCGACAGCTTGTGTGA
ACTGAAATCTAACAGAACAACAACAATAGTTTTTCAAGAGAGTTTGAAGATAATTAGTA
TTACTCAAA

>G1053 Amino Acid Sequence (domain in AA coordinates: 74-120)

MIPAEINGYFQYLSPEYNVINMPSSPTSSLNYLNDLIINNANNYSSSSNSQDLMISNNSTS
DEBHHQSIMVLDERKQRRMLSNRESARRSRMRKQRHLDLWSQVIRLRNENNCLIDKLN
RVSETQNCVLKENS KLKEEASDLRQLVCELKSNKNNNSFPREFEDN*

>G1137 (202..1248)

TACTTCAGACTTCTACTCAAACCAGTCACGTAGTTGGTTGGTGACATTTTCGCTGCATTTT
TCAATCTGTGATTGTTTTTCGTTTCGTTCTTTTACTATTTTCTCGAAAAGGACACAAG
AAGTATTGCATTCACTCAGTTGAGCAACTTAACAATCGTGTGTACTTTTTGAAGTTCCC
TTGAGCTAAACTGCTAAGAGCATGCCTCTGGATAAGAGGCAACGGGATTTGCCTCTGGGC
TTAAGTCCTCAAGCTTGCTTCAAGGATATAGTAGGTCGGTCTGTCTTCTAGAAATTCCT
CTCCCTGAGCTTGGGAACTATATGCAGCTAAGCTTCAGGCTCGTGTGTCAGCCACCA
CCATTCCAGTCTTTGCTGTGCAGTCATGATAAGGAGTCTTATGGAAAAAGATTCTCACGG
TCTGACATGCGGTCTTGGTGCGCTGCTGCTACTACTACTACTTCCACTTGGAGCATT
GAGTCTTCTCAGAAAAGACTTTTGATATTTCGATCAGTCAGGAGACCAGACTCGTCTATTA
CAATGTCCATTTCTCTACGGTTTCCATCTCATGCGGTGCAGAACAGTGAACTCTCT
GAGTTACAAGGTATAGAGAAAGCTTTCAAAGAAGATGGTGAAGAGTTTACAAGAGTGAT
GGAACAGAGTCAGAAATGCATGAAGACACTGAGGAGATCAATGCATTGCTATATTTCAGAT
GATGATTATGATGATGATTGCGAGAGTGATGATGAAGTAATGAGCACTGGTCACTCTCCT
TATCCAAATGAAGGAGTTTGCAACAAAAGGGAATTAGAAGAAATCGATGGTCTTGTAAA
AGGCAGAACTACTGGATAAGGTCAACAACATCAGCGACTTATCATCACTTGTGGGCACT
GAGAGCTCCACACAACCTCAATGGATCTTCTTTCTTAAGGACAAAAGCTCCCTGAATCA
AAAACCATATCGACCAAAGAGGACACTGGTTCTGGTCTGAGCAACGAGCAGTCGAAGAAA
GACAAGATCCGCACAGCTCTGAAAATACTCGAGAGCGTAGTCCCTGGTGCAAAGGAAAC
GAAGCGCTCTTACTTCTGGACGAAGCAATTGATTACCTAAAGTTGCTGAAACGAGACTTA
ATCTCCACAGAGGTTAAGAACCAAGCTCCACCACTCACAAGTCACCAATCTTGTGCTT
AAAGAGACAACATGGGGGAACAAGAAATCTGCAGACAGATAAGGCGTGAAAGATTCTGACG
AGTTAAAACGTGTGAAGTGGGTTTTTGGGTACGTATCCTTGCACCAGCTTT

>G1137 Amino Acid Sequence (domain in AA coordinates:264-314)

MPLDKRQRDLPLGLSPQACFKDIVGRSVLPRIPLPELGKLYAAKLQARCLQPPPFQSLLC
SHDKESYGRFRSRSDMRSCAAATTTTTPLGALESSQKRLLIQDQSGDQTRLLQCPFLR
FPSHAAEPVKLSELQIEKAFKEDGEEFHKSDGTESEMHEDTEINALLYSDDDYDDDC
ESDDEVMSGTGHPYPNEGVCNKRELEIDGPCKRQKLLDKVNNISDLSSLVGTESSTQLN
GSSFLKDKKLPESKTISTEKDTGSGLSNEQSKDKIRTALKILESVPVPAKGNEALLLLD
EADYLLKLLKRDLISTEVKNQSSSTTHKSPILLKKETTWTGRNLQTDKA*

>G1181 (113..1012)

CTCGATCTTTTAACCCCAATTATTACATATTACTCCTTCTACATTATCTTCTCTGCT
TTCGTGACTTTCAGGGGACACTTTTGTTTTTATAACTTACGCTTAAATCCCTATGAATTC
GCCGCCGTTGACGCAATGATTACCGGAGAATCATCGTCAAAAGATCTATCCCAACGCC
GTTTCTCAGAAAAACGTTTAACTCGTTGAAGATAGTTCATCGACGATGTTATCTCATG
GAACGAAGATGGTTCCTCTTTCATCGTATGGAATCCGACAGATTTGCTAAAGATTGCT
TCCTAAACACTTCAAACACAACATTTCTCTAGTTTCGTTTCGTCAGCTCAACACTTACGG
ATTCAAAAAAGTTGTACCGGATCGATGGGAGTTTCAAACGATTTCTTAAAGAGAGGAGA
AAAACGTCTTCTCCGTGAGATCCAACGTCCGAAAATAACAACGACGCATCAAACAGTTGT
TGCTCCTTCGTCGGAACAACGAAACAGACGATGGTGTATCACCGTCAAATTCGGGGGA
AGATAATAATAATAATCAGGTGATGTCTTCGTCTCCGTCTCGTGGTATTGTCAATCAAC
GAAGACGACTGGGAATGGTGGTTATCAGTGGAGTTATTGGAAGAGAACGAGAAGCTTCG
GAGTCAAAACATTACGCTAAACCGTGAGCTTACTCAGATGAAATCTATCTGCGATAATAT
CTATAGTCTCATGTGCAATTACGTCCGATCTCAGCCCACTGATCGGAGTTATTCTCCCGG
AGGTAGTAGTAGTCAACCGATGGAGTTTTTACCGGCGAAGCGTTTTTCGGAGATGGAGAT
TGAAGAAGAAGAAGAAGCGAGTCCGAGGTTGTTTGGTGTTCGGATTGGGTTAAACGGAC
GAGAAGTGAAGGTGTTCAAGTGAAGACGACGCGGTGGTTGGGAAAATTCCGATGAGGA
GACGCCGTGGTTGAGACATTATAATCGAACCAATCAGAGAGTTTGTAAATAAAAACGAAC
GGTTTAGATTGTGGTGTAGATATGTGCGGAAGTAGACGATTACAGCTTTTAAAGACAA
GCAGAGCACGTGCCCATCTGTTCAAGAAGTTTCTGCAATCTTGACTTCTTCTTTTAAAC
ACTTTGTGTTTTTTATTATTTAATAAACAATAAATGTTCTTTTTTTCAGTTTGTGTTTC
TTCAAAAATAGTTCCGCTGTTTCTAGACTTTCCTTTTTT

>G1181 Amino Acid Sequence (domain in AA coordinates: 24-114)

MNSPPVDAMITGESSSQRSIPTFLTKTFNLVEDSSIDDVISWNEDGSSFIWNPTDFAK
DLLPKHFKHNNFSSFVRQLNTYGFKKVVPDRWEFSNDFFKRGEKRLRLREIQRKITTTHQ
TVVAPSSEQRNQTMVVSNSGEDNNNNQVMSSSPSSWYCHQTKTTGNGGLSVELLEENE
KLRSQNIQLNRELTMKMSICDNIYSLMSNYVGSQPTDRSYSPGSSSQPMEFLPAKRFSE

MEIEEEEEASPRLFVPIGLKRTTRSEGVQVKTAVVGENSDEETPWL RHYNRTNQRVCN*

>G1228 (63..1139)

GCATTTATAATTACTCACTCATCTTCTTTTCATTACATTACATACCAAAACAAGAGCTCTC
AAATGGAAAGGTTTCAAGGACACATCAACCCCTGTTTCTTCGATCGAAAAACGGATGTGA
GAAGCCTCGAGGTTCAAGGATTTGCAGAGGCTCAAAGCTTTGCTTTCAAAGAAAAAGAGG
AAGAAAGCTTACAAGATACAGTTCCATTTCTACAGATGCTGCAAAGTGAAGACCCCTCAT
CGTTTTTTTCAATCAAAGAGCCAAACTTTCTGACGCTACTGTCTTCAAACCCCTCAAGG
AGCCTTGGGAACTCGAAAGATATCTTTCACTTGAGGATTACAAATTTCAATTCACCGGTCC
AATCTGAGACCAACCGCTTCATGGAAGGAGCCAATCAAGCTGTGTCAAGCCAAGAAATTC
CCTTTAGCCAAGCAACATGACACTCCCTTCTTCTACCTCATCACCCTCAGTGCACATT
CAAGACGAAAGCGCAAAATCAACCACTTGCTGCCTCAAGAAATGACTAGAGAAAAGAGAA
AGAGGAGGAAAACAAAACCAAGTAAAAACAATGAAGAGATTGAGAATCAAAGAATAAACC
ACATTGCTGTTGAACGAAACAGAAGACGTCAAATGAACGAACATATCAACTCTCTCCGGG
CCCTTCTCCACCTTCTTACATCCAACGAGGAGACCAAGCTTCCATAGTAGGAGGAGCAA
TAAACTACGTGAAGGTCCTCGAGCAAATCATACAATCTCTCGAATCGAAAAGAGAACGC
AACACAAAGTAACAGTGAGGTAGTAGAAAACGCACTTAATCATCTCTCAGGCATTTCTGT
CGAACGACCTGTGGACAACCTCTTGAAGATCAAACCTTGTATCCCCAAAATCGAAGCTACAG
TGATACAAAACCATGTGTCAGCCTTAAAGTTCAATGTGAGAAGAAACAAGGACAACCTTCTCA
AAGGAATCATATCACTTGAAAAGCTTAAACTCACTGTTCTTCTCATCTCAATATCACTACTT
CGTCTCATTCTCTGTTTCTTTATTCCTTCAACCTCAAGATGGAAGATGAGTGCGACTTAG
AGTCAGCCGACGAGATTACGGCGGCTGTTTCATCGGATTTTCGATATTCGACAATTTGAT
TAAACACATATAATTCCAAAATATTAACAGCTGACAAAATGGTATCTTTGCGGCC

>G1228 Amino Acid Sequence (domain in AA coordinates: 179-233)

MERFQGHINPCFFDRKPDVRSLEVQGFQFAEQSFQFKEKEEESLQDTPVFLQMLQSEDPSS
FFSIKEPNFLTLTSLQTLKEPWELELYLSLEDSQFHSPVQSETNRFMEGANQAVSSQEIP
FSQANMTLPSSTSSPLSAHSRRKRKINHLLPQEMTREKRKRRTKPSKNNEEIEENQRINH
IAVERNRRRQMNHEINSLRALLPPSYIQRGDQASIVGGAINYVKVLEQIIQSLESQKRTQ
QQNSEVVENALNHLSGISSNDLWTTLEDQTCIPKIEATVIQNHVSLKVQCEKKQGQLLK
GIIISLEKILKLTVLHLNITSSSHSVSYFNLKMEDECDLESADIEITAAVHRIFDIPTI*

>G1277 (51..512)

ATTCTAAAGTCCTCTCTCGGAAAGTAAGAGACTCAACTTCCGAGCCGCCATGGACGCCG
GAGTAGCAGTAAAGCTGACGTGGCAGTCAAAATGAAGAGAGAAAGACCATTCAAAGGGA
TCAGAAATGAGAAATGGGGGAAATGGGTTGCGGAGATTCGAGAACCCAACAAGCGTTCAA
GACTTTGGCTCGGCTCTTACTCTACTCCCAGAGCGGCGCGCGTGCATACGACACGGCTG
TCTTTTACCTCAGAGGACCACTGCTACGCTCAACTTCCCGGAGCTTCTGCCGTGTACCT
CCGCCGAGGATATGTGACGGCAACGATCAGGAAAAAGGCGACGGAGGTGGGAGCTCAAG
TAGATGCGATAGGGGCGACGGTGGTGCAGAACACAAACGCCGCCGCGCTTTTAGTCAAA
AGCGTGACTTTGGCGGCGGGTTATTAGAGCTTGTGACTTGAACAAGTTACCTGACCCGG
AAAATCTCGATGATGATTTGGTGGGAAAATAGACTGAAAAATAATAATAAATATCTTAC
AATGGTGGCTGTAGCTATCGTACGCGGAATGCTTGGGCTTGTGTTATATGACTACGTGGT
TACGGAAAGATTCTCTGTTTCGTCATTGTATTAAAAATTAATCCCACAAGTCAAACATA
CTGTACATTATTCTTAATTTAGTATTTCTTATTAATATCTATCATTTGTTTGGTGAACA
CCAGAAATATTAGACTATTAAATGTAACGAGTTTAAATATTTGATCATAATAACACCAAG
CTAGTTAAAGGTTAATATCTTGTACGAAGTCTTGAGTAAGTTCAATTGTATATATATG
TAACGGAAGAGGTTCTGTTCCGGTCCCAAGTGAAGTGATCAAAGGTGACTTCACATAAAA
AATAAAAAAAA

>G1277 Amino Acid Sequence (domain in AA coordinates: 18-85)

MDAGVAVKADVAVKMRERPFKIRMRKWGWVABIREPNKRSRLWLGSYSTPEAAARAY
DTAVFYLRGPTATLNFPELLPCTSAEDMSAATIRKKATEVGAQVDAIGATVVQNNKRRRV
FSQKRDFGGGLLELVDLNKLDPENLDDDLVGK*

>G1309 (53..859)

CGTCGACCTCTTAATTAAGACGACTTGAGAGAGAAAAGAAAGATACGTGGAAGATGACCAA
ATCTGGAGAGAGACCAAAACAGAGACAGAGGAAAGGGTTATGGTCACCTGAAGAAGACCA
GAAGCTCAAGAGTTTCTCTCTCTCGTGGCCATGCTTGCTGGACCACTGTTCCCATCTCT
AGCTGGATTGCAAAGGAATGGGAAAAGCTGCAGATTAAGGTGGATTAATTACCTAAGACC
AGGACTAAAGAGGGGGTCTGTTTAGTGAAGAAGAAGAAGAGACCATCTTGACTTTACATTC
TTCTTGGGTAACAAGTGGTCTCGGATTGCAAAATATTTACCGGGAAGAACAGACAACGA

GATTAAGAACTATTGGCATTCTATCTGAAGAAGAGATGGCTCAAATCTCAACCACAAC
CAAAAGCCAAATATCAGACCTCACAGAATCTCCTTCTTCACTACTTTCTTGC GGAAAAG
AAATCTGGAAACCGAAACCTTAGATCACGTGATCTCCTTCCAGAAATTTTCAGAGAATCC
AACTTCATCACCATCCAAAGAAAGCAACAACATGATCATGAACAACAGTAATAACTT
GCCTAAACTGTTCTTCTCTGAGTGGATCAGTTCTTCAAATCCACATCGATTACTCTC
TGCTTTTACAGATTCCAAGCACATTAATGAAACTCAAGATCAAATCAATGAAGAGGAAGT
GATGATGATCAATAACAACAACACTACTCTTCACTTGAGGATGTCATGCTCCGTACAGATTT
TTTGCAGCCTGATCATGAATATGCAAATTATTATTCTTCTGAGATTTCTTCATCAACAG
TGACCAAATATGTCTAAGAAGAGTGAATATGATCGTAAGAGGAACATAAGCTAGTTAC
TTGTGTTACAGC

>G1309 Amino Acid Sequence (domain in AA coordinates: 9-114)
MTKSGERPQRQRKGLWSPEEDQKLKSFILSRGHACWTTVPILAGLQRNGKSCRLRWINY
LRPGLKRGFSFSEEEETILTLHSSLGNKWSRIAKYLPGRTDNEIKNYWHSYLLKRWLKSQ
PQLKSQISDLTSPSSLLSCGKRNLLETLDHVISFQKFSENPTSSPSKESNNNMIMNNS
NNLPKLFFSEWISSNPHIDYSSAFTDSKHINETQDQINEEVMMINNNNYSSLEDVMLR
TDFLQPDHEYANYSSGDFFINSDQNYV*

>G1314 (1..990)
ATGGGAAGAGCTCCGTGTTGCGACAAGACAAAAGTGAAGCGAGGGCCTTGGTCGCCTGAA
GAAGACTCTAAACTTAGAGATTACATTGAAAAGTATGGTAATGGTGGAAATTGGATCTCT
TTCCCCCTCAAAGCCGGTTTGAGGAGATGTGGGAAGAGTTGTAGACTGAGGTGGCTAAAC
TATTTGAGACCAACACATAAAGCATGGTGACTTCTCTGAGGAAGAAGACAGGATCATTTTT
AGTCTCTTCGCTGCCATAGGAAGCAGGTGGTCAATAATAGCAGCTCATCTACCGGGACGA
ACAGACAACGACATAAAAACTATTGGAACACAAAGCTAAGGAAGAACTCTTGCTTTCT
TCCTCTGATTTCATCATCATCAGCCATGGCTTCTCCTTATCTAAACCTATTTCTCAGGAT
GTGAAAAGACCAACCTCACCACAAACAATCCCATCTTCTTCTTACAATCCGTATGCTGAA
AACCCTAATCAATACCCAACAAAATCCCTCATCTCCAGCATCAATGGCTTCGAAGCTGGT
GACAAACAGATAATTTCTATATTAACCCTAATTATCCTCAAGATCTCTATCTCTCGGAC
AGCAACAACAACACCTCGAACGCAATGGTTTTCTTGCTCAACCACAATATGTGTGATCAG
TACAAGAACCACACAGTTTTTCTTTCAGACGTCATGAGGGATAAGATCAGAGATTATGATG
AAGCAAGAAGAGATAATGATGATGATGATGATAGACCACCATGACCAGAGGACAAAA
GGGTACAATGGGGAATTCACACAAGGGTATTATAATTACTACAATGGGCATGGGGATTTG
AAGCAAATGATTAGTGAACAGGCATAATTCTAACATAAACATGGGTGGTTTCAGGTTCA
TCTTCTAGTTGATAAGCAACCTAGCTGAGAACAAAAGCAGTGGTAGCCTCCTACTAGAA
TACAAATGCTTGCCCTATTTCTACTCCTAG

>G1314 Amino Acid Sequence (domain in AA coordinates: 14-116)
MGRAPCCDKTKVKRGPSPEEDSKLRDYIEKYNGGNWISFPLKAGLRRCGKSCRLRWLN
YLRPNIKHGFSEEDRIIFSLFAIGSRWSIIAAHLPGRTDNDIKNYWNTKLRKKLLSS
SSDSSSSAMASPYLNPISQDVKRPTSPPTIPSSSYNPAENPNQYPTKSLISSINGFEAG
DKQIISYINPNYPQDLYLSDSNNTSNANGFLLNHNMCQYKNHTSFSSDVNGIRSEIMM
KQEEIMMMMHIDHIDQRTKGYNGEFTQGYNNYNGHGLDKQMISGTGTSNINMGGSGS
SSSSISNLAENKSSGSLLEKCLPYFYS*

>G1317 (1..849)
ATGGGAAGATCACCTTGTTGTGATAAAAATGGAGTGAAGAAGGGACCATGGACTGCTGAG
GAGGATCAGAACTCATCGATTATTCGATTTTCATGGTCCTGGCAATTGGCGTACGCTC
CCCAAAAATGCTGGACTCCATAGATGTGGAAAAAGCTGCCGTCTTCGATGGACCAATTAT
CTAAGACCGGACATCAAGAGAGGAAGATTCTCGTTCGAGGAAGAAGAACTATCATTGAG
CTACACAGTGTTATGGGAAACAAGTGGTCAGCAATAGCCGCTCGTCTACCAGGGAGGACC
GATAACGAAATAAAAAACCATTTGGAACACTCACATCCGCAAGAGACTTGTAAAGGAGTGGT
ATCGACCCTGTTACTCATTCTCCACGCCTTGATCTTCTTGATTTGTCTCTCACTTTTGAGT
GCACTTTTCAACCAGCAAACTTTTCAGCAGTTGCAACACATGCGTCTTCTCTTCTTAAT
CCTGATGTATTGAGGTTGGCCTCTCTACTACTGCCACTTCAAACCCCTAATCCAGTTTAC
CCATCGAACCTCGACCAAAATCTTCAAACCTCAAATACATCATCAGAATCGTCTCAACCA
CAAGCTGAGACTAGTACAGTCCCAACAACTATGAACTTCATCATTGGAGCCTATGAAC
GCAAGACTCGACGACGTTGGTCTTGCAGATGTATTACCACCTTTGTGAGAGAGTTTGGAC
TTAGACTCGCTCATGTCAACGCCAATGTCTTCTCCACGACAAAATAGCATTGAAGCAGAA
ACCAACTCCAGCACTTTCTTCGACTTTGGAATTCGGAAGATTTTCATCTTAGATGACTTT
ATGTTTTAA

>G1317 Amino Acid Sequence (conserved domain in AA coordinates:13-118)

MGRSPCCDKNGVKGPWTAEDQKLIDYIRFHGPGNWRTLPKNAGLHRCGKSCRLRWTNY
LRPDIKGRGRFSFEEETIIQLHSVGMGNKWSAIAARLPGRTDNEIKNHNTHIRKRLVRSG
IDPVTHSPRLDLLDLSSLLSALFNQPNFSAVATHASSLLNPDVLRLLASLLPLQNPVY
PSNLQNLQTPNTSSSESSQPAETSTVPTNYETSSLEPMNARLDDVGLADVLPPLSESF
LDSLSTPMSSPRQNSIEAETNSSTFFDFGIPEDFILDDFMF*

>G1323 (49..870)

AAGAGGGAATCTCAAAAGTGTGTGTCTGTGAGAGAGGAGAGAGAGAAATATGGGCAAAGGA
AGAGCACCATGTTGTGACAAAACCAAGTGAAGAGAGGACCATGGAGCCATGATGAAGAC
TTGAAACTCATCTCTTTCAATCACAAGAATGGTCATGAGAATTGGAGATCTCTCCCAAAG
CAAGCTGGATTGTTGAGGTGTGGCAAGAGTTGTCGTCTGCGATGGATTAATTACCTCAGA
CCTGATGTGAAACGTGGCAATTTTCAGTGCAGAGGAAGAAGACACCATCATCAAAC TTCAC
CAGAGCTTTTGGTAACAAGTGGTGAAGATTGCTTCTAAGCTGCCTGGAAGAACAGACAAT
GAGATCAAGAATGTGTGGCATAACATCTCAAGAAAAGATTGAGCTCGGAAACTAACCTT
AATGCCGATGAAGCGGGTTCAAAGGTTCTTTGAATGAAGAAGAGAACTCTCAAGAGTCA
TCTCCAAATGCTTCAATGTCTTTTGCTGGTTCCAACATTTCAAGCAAAGACGATGATGCA
CAGATAAGTCAAATGTTTGAGCACATTTCTAAGCTTATAGCGAGTTTACGGGGATGTTACAA
GAGGTAGACAAACCAGAGCTGCTGGAGATGCCCTTTTGATTTAGATCCTGACATTTGGAGT
TTCATAGATGGTTTCAGATCATTTCCAACAACAGAGAAACAGAGCTCTTCAAGAGTCTGAA
GAAGATGAAGTTGATAAATGTTTAAAGCACCTGGAAGCGAACTCGGGTTAGAAGAAAAC
GATAACCAACAACAACAACAGCATAAACAGGGAACAGAAGATGAACATTCATCATCACTC
TTGGAGAGTTACGAGCTCCTCATACATTAATGAAGCCATAAAGCAAGTCATTTTACCTT
GAAAATGGAATTATTAGCTAACTTATTGGCATTATTAGTATATAAGCAAGATCAGATAGG
CGCATGTAGTAGCAACAACGAAGAAACGTCGAATTGTAGACAAAATGTAGATATTACAGA
GTTGAAAGATTGTATTTTGCAAATGATTGCTTTGTAGTGAAATCAAGTTATCACAAAAA
AAAAAAA

>G1323 Amino Acid Sequence (domain in AA coordinates: 15-116)

MGKGRAPCCDKTKVKGPSHDEDLKLISFIHKNHGNWRS LPKQAGLLRCGKSCRLRWI
NYLRPDVKRGNFSAEEEDTIIKLHQSFGNKWSKIASKLPGRTDNEIKNVWHTLKKRLSS
ETNLNADEAGSKGSLNEEENSQESSPNASMSFAGSNISSKDDDAQISQMFHILTYSEFT
GMLQEVDPKPELLEMPFDLPDIWSFIDGSDSFQQPENRALQESEDEVDKWFKHLESELG
LEENDNQQQQHKQGTEDHSSSSLESYELLIH*

>G1332 (1..606)

ATGGAATGCAAAAGAGAAGAAGGGAAGTCTTACGTGAAGAGAGGGTGTGGAAACCAGAA
GAAGATATGATATAAAAGCTATGTTGAGACTCATGGTGAAGGAAACTGGGCAGACATT
TCTCGTAGATCCGGGTTGAAGAGAGGAGGAAAAGCTGTAGGCTGAGATGGAAGAACTAT
CTAAGACCAAATATCAAAAGAGGAAGCATGTCACCACAAGAACAAGACCTTATCATCCGC
ATGCATAAGCTTCTTGAAACAGATGGTCGTTGATCGCTGGTCGCCTTCCAGGTGCTACT
GACAATGAAGTGAAGAACTACTGGAATACTCATTTGAACAAGAAACCTAATTCCCAGAAA
CAGAATGCACCTGAATCAATCGTCGGCGCCACTCCTTTCAAGGATAAGCCAGTTATGTCT
ACAGAACTGAGAAGAAGCCATGGAGAAGGAGGAGAAGAGGAGAGCAATACCTGGATGGAG
GAGACCAACCACTTTGGCTATGACGTCCACGTAGGATCTCCCTTGCCACTTATTTCCAC
TACCCAGACAACACTCTCGTGTGTGACCCATGTTTTCCTTTACCGATTTCTTTCTCTG
CTTTAG

>G1332 Amino Acid Sequence (conserved domain in AA coordinates:13-116)

MECKREEGKSYVKRGLWKPEEDMILKSYVETHGEGNWADISRRSGLKRGKSCRLRWKNY
LRPNIKRGSMSPEQDLIIRMHKLLGNRWSLIAGRLPGRTDNEVKNYWNTHLNKKPNSRK
QNAPEISIVGATPFTTBKPVMS TELRRSHGEGGEESNTWMEETNHFGYDVHVGSPPLPLISH
YPDNTLVFDFPCFSFTDFFPLL*

>G1334 (76..885)

ATAGCTCCCAACTAATAGGAATCTCAAGCTTCTCACTCTCTCTTGTTTTCCATTGGACT
TTTGGAAACATAAGCTATGCAAACTGAGGAGCTTTTGTGCGCACCAACAGACTCCTTGGTGG
AATGCTTTTGGATCTCAGCCGTTGACTACAGAGAGCCTTTCCGGCGAAGCTTCTGATTCA
TTCACCGGAGTTAAGGCAGTTACTACGGAGGCAGAACAAAGGTGTGGTGGATAAACAACT
TCTACAACTCTCTTCACTTCTCACCTGGTGGTGAAGAGATTCAAGAGATGTGCCAAAG
CCTCATGTTGCTTTTCGCGATGCAATCAGCTTGCTTCGAGTTTGGATTGCTCAGCCAATG
ATGTACACAAAGCATCCTCATGTTGAACAATACTATGGAGTTGTTTCAGCATACGGATCT

CAGAGGTCTTCGGGCCGAGTAATGATTCCACTGAAGATGGAGACAGAAGAAGATGGTACC
ATCTATGTGAACTCAAAGCAGTACCATGGAATTATCAGGCGACGCCAGTCCCGAGCAAAG
GCTGAAAACTGAGTAGATGCCGTAAGCCATATATGCATCACTCAGCCATCTCCATGCT
ATGCGCCGTCCTAGAGGATCTGGCGGGCGTTTCTTGAACACCAAGACAGCTGATGCGGCT
AAGCAGTCTAAGCCGAGTAATTCTCAGAGTTCTGAAGTCTTTCATCCGGAATGAGACC
ATAAACTCATCGAGGGAAGCAAATGAGTCAAATCTCTCGGATTCTGCAGTTACAAGTATG
GATTACTTTCTAAGTTCGTCGGCTTATTCTCTGGTGGCATGGTCATGCCATCAAGTGG
AATGCAGCAGCAATGGATATTGGCTGCTGCAAACTTAATATATGATCAGCAGATAGGGGA
CAAGACATGATTGGTCAACAGTCCTTTGTCTTGTCCCTTATCTTTCAGCCAAACGGAAA
GAGAACTTGTGTCTTGGAAAAAAGACATTGAGTTTCCCTTGGTTTATAAGATTGGTCCTTT
TACCATCCGTTTGGCTGTAAACAGGCAAATCATCTTGGCTCATGCTTCATCAAGTTCTT
ATCTTCGCTCTGTTTTCTTCTACGCATCTTCATAAGATCTCTGAAGTGAATAACATTT
CCTAGCATCATGTTTCAACTAGTGTGTGTGTAAGAACTCTGCCTTATTTCCAGATGAT
GTATTGTGTGTAACGTGTTTATGAAACAAACGTAAGACTTTCAGTTAAAAA
AAAAAAAAAAAAA

>G1334 Amino Acid Sequence (domain in AA coordinates: 18-190)
MQTELLSPQPWPWNFAFGSQPLTTESLSGEASDSFTGVKAVTTEAEQGVVDKQSTTLF
TFSPGGEKSSRDVPHVAFAMQSACFEFGAQPMMYTKHPHVEQYYGVVSAYGSQRSSG
RVMIPKMETEEDGTIYVNSKQYHGIIRRRQSRKAELSRCKPYMHHSRHLHAMRRPR
GSGGRFLNKTADAQSKPSNSQSSEVFHPENETINSSREANESNLSDSAVTSMDYFLS
SSAYSPGGMVMPKWNAAAMDIGCKLNI*

>G1381 (32..802)
CAGCTTTAACTACTCTCTCTCTCTCAATGGGAAAACAAATCAACATAGAGAGTAG
TGCTACTCATCATCAAGACAATATTGTTTCCGTTATAACAGCCACGATATCCTCCTCCTC
CGTCGTAACGTCTTCGTGAGACTCTTGGTCTACCTCCAAAAGATCGTTAGTGCAAGACAA
TGACTCCGGAGGGAAACGGCGGAAGAGCAACGTTAGTGATGATAACAAGAATCCGACGTC
GTATAGAGGAGTGAGGATGAGGAGTTGGGGAAAATGGGTGTCGGAGATTAGAGAGCCGAG
GAAGAAATCAAGAATATGGCTTGGCACTTATCCAACGGCAGAGATGGCAGCTCGTGCTCA
TGATGTGGCGGCTTTAGCTATTAAAGGCAACTCCGGTTTTCTTAATTTCCCTGAATTATC
CGGTTTGCTTCTCGTCCGTTAGCTGCTCTCCTAAGGATATACAAGCTGCAGCTACCAA
AGCCCGCAAGCAACCACGTGGCACAACCCGTTATCGATAAGAAATTAGCTGATGAGCT
AAGCCACTCTGAGTTGTTGTCTACCGCTCAGTCTTCGACTTCTAGTAGTTTCGTGTTTTCT
TTCGGACACGTCCGAGACTTCTAGTACGGACAAGGAAAGCAACGAAGAGACGGTGTGTTGA
TTTGCCGGACCTTTTACCGGACGGGCTTATGAACCCAAACGATGCGTTTTGTTTATGCAA
CGGCACCTTTACGTGGCAGCTTACGGAGAGGAGGATGTAGGGTTCAGGTTTGAAGAGCC
GTTTAATTGGCAAAATGACTAAACCGCCCTCCACTTGCTTACTGTAATTACTAACATATA
ATTTTCTTGATAAAGAACATATATTCCATTACGGTATTAACTAATCTTTTCTATCCTTT
TCTCTTTTCTGTTTCTACATCTGAGTATATTGTCACTATGTGAAAAAATTGATCTCGTT
TTGAATATTTACTTTTCAAATTTGAAGTAACGCAAGTGATTGATAAAAAAAAAAAAAA
>G1381 Amino Acid Sequence (domain in AA coordinates: TBD)
MGKQINIESSATHHQDNIVSVITATISSSSVVTSSSDSWSTSKRSLVQDNDSSGKRRKSN
VSDDNKNPTSIRGVRMRWSGKWVSEIREPRKKSRIWLGYPTAEMAARAHDAALAIKGN
SGFLNFPPELSGLLPRPVSCSPKDIQAAATKAAEATTWHKPVIDKKLADELSHSELLSTAQ
SSTSSSFVFSSDTSETSDKESNEETVFDLPDLFTDGLMNPNDAFCLCNGTFTWQLYGE
EDVGFRFEEFPFNQND*

>G1382 (90..1763)
CTCTCATTTCCCATAGCTGAGAGCTTCTTCTACTTTCCCTTAGCTTCTTTTTCTTCA
TTTTTGTCTACCCFTGCGAATCTCTGAAATGAACCTCAAGCTAATGACCGGAAGGAGT
TTCAGGGAGATTGTTTCGGCGACGGGAGATCTCACGGCAAAGCACGATTCAGCTGGAGGAA
ACGGAGGTGGAGGTGCTAGGTATAAGCTGATGTACCCGGCCAAGCTTCCGATCTCGAGGT
CGACTGATATCACGATTCCTCTGGGTTGAGTCCGACTTCGTTTTTGAATCTCCTGTTT
TCATCTCCAACATCAAGCCAGAACCTTCCCCTACTACTGGTCTTTGTTCAAGCCTCGAC
CAGTGACATTTCTGCTAGCTCAAGTTCTTATACAGGCAGGGGTTCCATCAGAACACCT
TTACTGAGCAGGAAGTCCAGTGAATTTGAGTTCAAGCTCCTGCATCAAATATGGTATATG
CAGAGCTTGGCAAGATTAGAAGTGAGCCACCAGTACATTTTCAAGGCCAGGGCCATGGAT
CCTCACACTACCTTCTTCGATCAGTGATGCTGCAGGTTCTCAAGTGAGCTAAGCCGGC
CAACTCCTCTTGTGATGACACCAACGAGCTCAGATATTCCGGCTGGATCTGATCAAG

AGGAATCAATCCAGACTTCCCAAAATGACTCCAGAGGAAGCACTCCATCCATCTTGGCTG
ATGATGGTTATAACTGGAGAAAATATGGTCAAAAGCATGTCAAAGGGAGTGAATTTCCCC
GGAGCTATTATAAATGTACACATCCTAATTGTGAAGTGAAAAAGTTATTTGAAAGATCTC
ATGATGGGCAGATACCGGATATTATATACAAGGGTACACATGACCATCTTAAACCTCAAC
CTGGTCGCCGAAACTCTGGTGGTATGGCTGCACAAGAAGAAAGGCTAGACAAGTATCCTT
CTTCAACTGGCCGAGATGAGAAGGGATCTGGCGTCTACAACCTGTCTAACCCCAATGAAC
AAACTGGTAACCTGAAGTACCTCTATCTCAGCATCTGACGATGGTGGAGAAGCGGCAG
CGTCAAATAGGAATAAAGATGAGCCGGACGATGATGATCCATTCTCAAAACGGAGGAGGA
TGGAGGGTGGATGGAAATAACTCCACTAGTGAAACCCATCCGGGAGCCTCGGGTTGTG
TTCAAACCTCTGAGTGAGTTGACATTCTGGATGATGGTTATAGATGGCGCAAATATGGGC
AGAAAGTCGTAAGGGGGAACCCAAATCCCAGGAGCTACTACAAATGCACAGCTCATGGAT
GCCAGTGAAGAAAACACGTGGAGAGAGCATCACATGATCCAAAAGCTGTAATAACAACAT
ACGAAGGCAAACACGATCATGATGTTCCCACTTCAAAGCTAGCAGCAATCACGAAATCC
AGCCTCGGTTACAGACGATGAAACAGACACCATCAGCCTCAATCTTGGTGTGGAATCT
CATCTGATGGACCTAACACGCTTCCAACGAACATCAGCACCAGAATCAACAACCTGTCA
ACCAAACCTACCCAAATGGAGTCAATTTTCAAGTTTGTTCATGCTAGTCCCATGTCTCCT
ACTATGCTAGCTTAAATAGCGGTATGAATCAGTACGGCCAGAGAGAAACAAAGAACGAGA
CTCAAATGGTGACATCTCGTCTTGAACAATTATCTTACCCATATCCGCCAACATGG
GGAGAGTACAAATCGGGTCCGTAAACAAAAAGTAAGCAACATTATGTACGGGATCTTCTT
AGGTAGGAATGGGACGAGGCTTGTCTATATAATTCTTATTTCTTACAGAGAGCTGA
TCTTGATTCAAACCTATCTCCACCATATATATTTGTTTGTGTACCTGTATTGAGTTCCAA
AAATGTTATGTAATAACACAACAAGATGTTAATGCTTTTATTTAAACAAGAAACAGCA
ATATTACTACAAAAA

>G1382 Amino Acid Sequence (domain in AA coordinates: 210-266, 385-437)

MNPQANDRKEFQGD SATGDLTAKHDSAGNGGGGARYKLMSPAKLPISRSTDITIPPGL
SPTSFLESPVFISNIKPEPSPPTGSLFKPRPVHISASSSYTGRGFHQNTFTEQKSSEFE
FRPPASNMVYAE LGKIRSEPPVHFQGGHGHSSHPSSISDAAGSSSEL SRPTPPCQMTPT
SSDIPAGSDQESIQT SQNDSRGSTPSILADDGYNWRKYGQKHVKGSEFPRSYKCTHPN
CEVKKLFERSHDQITDIIYKGT HDHPKPQGRNRSGGMAAQEERLDKYPSS TGRDEKGS
GVYNLSNPNEQTGNPEVPPISASDDGGEAAASNRNKDEPDDDDPFSKRRRMEGAMEITPL
VKPIREPRVVVQTLSEVDILDDGYRWRKYGQKVVRGNPNPRSYKCTAHGCPVRKHVERA
SHDPKAVITTYEGKHDHVDPTSKSSSNHEIQPRFRPDEDTISLNLGVGISSDGPNHASN
EHQHQNQQLVNQTHPNGVNF RFVHASPMSSYYASLNSGMNQYQRET KNETQNGDISSLN
NSSYPYPNMGVRVQSGP*

>G1435 (8..904)

GTGAAACATGGGGAAGGAAGTTATGGTGAGCGATTACGGTGACGACGACGGAGAAGACGC
CGGCGGCGGCGATGAATATAGGATTCCGGAATGGGAAATTGGTTTACCCAACGGAGATGA
TTTGACTCCGTTATCTCAATATCTAGTCCCGTCGATTCTCGCGTTAGCTTTCAGCATGAT
CCCAGAACGAAGCCGTACAATTCACGACGTCAATCGCGCGTCGCAAATCACGCTCTCTTC
GTTGAGAAGCAGTACCAATGCTTCGTCTGTGATGGAGGAGGTCTGGATCGAGTTGAATC
GAGTGTTCAGGATCAGATCCGAAGAAACAGAAGAAATCGGATGGTGGTGAAGCAGCGGC
GGTGGAGGATTCCACGGCGGAGGAAGGAGACTCCGGGCCTGAAGACGCGTCTGGGAAGAC
ATCGAAACGACCGGCTTTAGTGTGGACACCGCAGCTACACAAGAGATTGTGGACGTTGT
GGCTCATCTAGGGATTAAAAACGAGTGCCGAAGACGATTATGCAGCTGATGAACGTGGA
AGGACTTACTCGTGAGAACGTTGCGTCTCATTTGCAGAAATATAGGCTTTACCTTAAACG
GATTCAAGGATTGACGACGGAAGAAGATCCTTATTCGTCGTCGGATCAGCTCTTCTTTC
AACGCCGGTTCTCCACAGAGCTTCAAGACGGCGGAGGAAGTAACGGAAGTTGGGGGT
TCCGGTTCCGGTTCGTCGATGGTGCTTATTCAGGCTATGGGAATCAAATGGGTATGCA
AGGATATTATCAACAGTATAGTAACCATGGCAATGAATCAAACCAATATATGATGCAGCA
GAATAAGTTTGAACAATGGTGACATATCCTTCTGTTGGTGGTGGTGACGTGAATGACAA
GTAAATGGATCTTAAAGGCTTATAATTTGCTCTACAGAGAGATACTGGTTCTTGGCTTAT
GGTTTATTTTCCCACTTCATGAGGTGTTGTGACTTTTAATTCTCCATGTTTTCACACA
AGTCTTTATTGCTTTGTATAGAAAATGATTTTCAGAGAAATCACTGGGAAGCTTGGTATT
GTTGGAGGATGAAGCCTTCTATGAATGATTTAGTTTCTTACTGTCTCCATTCTTTATGAG
GTAATAAAGCCTTCTTTTGTCTCATCGCTTGTAGTCTTCTTAAATTCAAGACAGCGTCACA
TGTTTGTTCGGTTATGTTAATTGTTCTTTCTTTGGATAATGAAGATAGCATCAGGTCTC
ATGTCTCCTCACTTTGATAAA

>G1435 Amino Acid Sequence (domain in AA coordinates: 146-194)
MGKEVMVSDYGDGDDGDEAGGGDEYRIPEWEIGLPNGDDLTPLSQYLVP SILALAFSMIPE
RSRTIHDVNRASQITLSSLRSSSTNASSVMEVVDREVSSVPGSDPKKQKSDGGEAAAVE
DSTAEEGDSGPEDASGKTSKRPRLVWTPQLHKRFVDVVAHLGIKNAVPKTIMQLMNV EGL
TRENVAHLQKYRLYLKRIQGLTTEEDPYSSSDQLFSSTPVPPQSFQDGGGSNGKLGVPV
PVPSMVPPIPGYGNQMGMQGYQYQYSNHGNESNQYMMQONKFGTMVTYPSVGGGDVNDK*

>G1537 (1..783)

ATGGAAAACGAAGTAAACGCAGGAACAGCAAGCAGTTCAAGATGGAACCCAACGAAAGAT
CAGATCACGCTACTGGAATCTTTTACAAGGAAGGAATACGAACCTCCGAGCGCCGATCAG
ATTACAGCAGATCACCGGTAGGCTTCGTGCGTACGGCCATATCGAAGGTAAAAACGTCTTT
TACTGGTTCAGAACCATAGGCTAGGCAACGCCAAAAGCAGAAACAGGAGCGCATGGCT
TACTTCAATCGCCTCTCCACAAAACCTCCCGTTTCTTCTACCCCCCTCCTTGCTCAAAC
GTGGGTGTGTGTGTCAGTCCGTACTATTTACAGCAAGCAAGTGATCATCATATGAATCAACAT
GGAAGTGATACACAAACGATCTTCTTCACAGAAACAATGTGATGATTCCAAGTGGTGGC
TACGAGAAACGGACAGTCACACAACATCAGAAACAACCTTTCAGACATAAGAACAACAGCA
GCCACAAGAATGCCAATTTCTCCGAGTTCAGTACAGATTGACAGATTGCCCTCCGTGAT
AACTGTTATGCCCGGTGAGGACATTAACGTCAATTCCAGTGGACGGAAAAACACTCCCTCTT
TTCTCTCTTCAGCCTTTGAATGCAAGTAATGCTGATGGTATGGGAAGTTCAGTTTGGC
CTTGGTAGTGATTCTCCGTGGATTGTTCTAGCGATGGAGCCGGCCGAGAGCAGCCGTTT
ATTGATTTCTTTCTGGTGGTTCTACTTCTACTCGTTTCGATAGTAATGGTAATGGGTTG
TAA

>G1537 Amino Acid Sequence (domain in AA coordinates: 14-74)
MENEVNAGTASSSRWNPTKDQITLLENLYKEGIRTPSADQIQITGRRLRAYGHIEGKNVF
YWFQNHKARQRQKQKQERMAYFNRLHKTSTRFFYPSPCSNVGCVSPYYLQQASDHMNQH
GSVYTNDLLHRNNVMIPSGGYEKRTVTQHQQQLSDIRTTAATRMPISPSSLRDFRFRALRD
NCYAGEDINVNSSGRKTLPLFPLQPLNASNADGMGSSSFALGSDSPVDCSSDGAGREQPF
IDFFSGGSTSTRFDSNNGNL*

>G1545 (67..729)

CATCACCAATCTTTTGAATCTAAGAGAGAGAAGAAGAAGAGGTCTAGAGAACGAAAAGA
AGAAACATGAATAACCAGAATGTAGATGATCATAATCTTCTACTCATTTCTCAATTGTAC
CCTAATGTCTATACTCCATTAGTACCACAACAAGGAGGAGAAGCAAAACCAACACGGCGG
AGGAAAAGGAAGAGCAAGAGTGTGTGGTGGCAGAGGAGGGTGAAAACGAAGGCAATGGG
TGGTTTTAGAAAGAGAAAATTGAGTGATGAGCAAGTAAGAATGTTGGAGATTAGCTTTGAA
GACGATCATAAGCTTGAATCCGAGAGGAAAGATCGGCTTGCTTCTGAGTTAGGGCTTGAT
CCTCGTCAAGTCGCCGTCTGGTTCCAAAACCGCCGTGCACGGTGGAAGAACAACAGAGTC
GAGGATGAATACACTAACTCAAGAATGCATACGAAACCACCGTCGTTGAGAAATGTCGT
CTTGATTCTGAGGTTATTACCTAAAGGAACAACCTTACGAGGCTGAAAGAGAGATCCAA
CGGCTTGCAAAAAGAGTTGAAGGAACCTTAAAGTAACAGTCCTATCTCATCCTCTGTGACC
ATTGAAGCCAATCATACGACACCGTTTTTTGGAGATTACGACATCGGATTGACGGTGAG
GCTGACGAGAACTTGCTCTACTCGCCAGATTACATTGATGGATTAGACTGGATGAGCCAA
TTTATGTAAAAAATAAGCTAATCTATTTTCAGTCGTAGTATAG

>G1545 Amino Acid Sequence (domain in AA coordinates: 54-117)
MNNQNVDDHNNLLISQLYPNVYTPLVPQQGGEAKPTRRRKRKSKSVVVAEEGENEGNGWF
RKRKLSDEQVRMLEISFEDDHKLESERKDRLASLGLDPRQVAVWFQNRARRWKNKRVED
EYTKLKNAYETTVVEKCRLDSEVIHLKEQLYEAEREIQRLAKRVEGTLSNSPISSSVTIE
ANHPTTFFGDYDIGFDGEADENLLYSPDYIDGLDWM SQFM*

>G1641 (1..867)

ATGGAGGTTATGAGACCGTCGACGTACACGTGTCAGGTGGGAACCTGGCTCATGGAGGAA
ACTAAGAGCGCGCTCGCAGCTTCTGGTGAAGGTGCCACGTGGACGGCGGCAGAGAACAAG
GCATTCGAGAATGCTTTGGCGGTTTACGACGACAACACTCCTGATCGGTGGCAGAAGGTG
GCTGCGGTGATTCCGGGGAAGACAGTGAGTGACGTAATTAGACAGTATAACGATTTGGAA
GCTGATGTCAGCAGCATCGAGCCGGTTAATCCCGGTCCCCGGTTACATCACCTCGCCG
CCTTTACTCTAGATTGGGCGGCGGCGGTGGCGGATGTAACGGGTTTAAACGGGTCAT
CAGGTTTGTAATAAACGGTCGACGGCCGGTAGATCGCCGAGCTGGAGCGGAAGAAAGGC
GTTCTTGGACGGAGGAAGAACAAGCTATTTCTAATGGGTTTGAAGAAATATGGGAAA
GGAGATTGGAGAAACATATCTCGGAACCTTTGTGATAACCGCAACGCCAACACAAGTAGCT
AGCCACGCCCAAGAGTACTTCATCCGGCAACTTTCCGGCGGCAAGGACAAGAGACGAGCA

AGCATTCACGACATAACCACCGTAAATCTCGAAGAGGAGGCTTCTTTGGAGACCAATAAG
AGCTCCATTGTTGTTGGAGATCAGCGTTCAAGGCTAACCGCGTTTCTTGGAAACCAAACG
GACAACAATGGAACACAGGCAGACGCTTCAATATAACGATTGGAAACGCTATTAGTGGC
GTTCAATTCATACGGCCAGGTTATGATTGGAGGGTATAACAATGCAGATTCTTGCTATGAC
GCCCCAAACACAAATGTTTCAACTATAG

>G1641 Amino Acid Sequence (domain in AA coordinates: 139-200)
MEVMRPSTSHVSGGNWLMEETKSGVAASGEGATWTAENKAFENALAVYDDNTPDRWQKV
AAVIPGKTVSDVIRQYNDLEADVSSIEAGLIPVPGYITSPFPLDWAGGGGGCNGFKPGH
QVCNKRQAGRSPELERKKGVPTWTEEEHKLFLMGLKKYKGDWRNISRNFVITRPTQVA
SHAQKYFIRQLSGGKDKRRASIHDIITVNLEEEASLETNKSSIVVGDQSRSLTAFPNQT
DNNGTQADAFNITIGNAISGVHSYGQVMIGGYNNADSCYDAQNTMFQL*

>G165 (19..699)

CTTCAAAACATCTAAAAAATGGTGAAAAAACTCTTGGTCGTAGAAAGGTAGAGATAGTG
AAAATGACTAAGGAATCAAACCTTCAAGTCACATTTTCCAAGAGAAAAGCTGGTCTTTT
AAGAAGGCTAGTGAATTTTGCACATTATGTGATGCAAAAATTGCGATGATCGTGTTC
CCAGCTGGAAGATTTTCTTTTGGTCATCCAAATGTTGATGTTCTGCTTGACCACTTT
CGAGGGTGTGTTGTAGGACACAACAACACAAACCTTGATGAAAGCTACACAAAGCTCAT
GTTCAAATGCTCAACAAATCCTACACTGAGGTGAAGGCGGAAGTAGAAAAAGAACAAG
AATAAGCAGTCGCGGGCTCAAATGAAAGAGAAAACGAAAACGCTGAGGAGTGGTGGAGT
AAGTCTCCATTAGAACTCAACTTAAGTCAATCAACCTGTATGATACGTGTTCTTAAAGAT
TTGAAGAAGATAGTTGATGAAAAAGCAATTCAATTAATCCATCAAACAAACCCAACTTC
TATGTTGGAAGTTCTAGCAATGCTGCTGCTCCAGCAACTGTTAGTGGTGGTAATATCTCC
ACAAACCAGGGGTTCTTTGATCAAAACGGAATGACGACTAATCCTACTCAAACACTTCTG
TTTGGATTGATATTATGAATCGCACACCAGGAGTTTAAATAAGTCTATCCTCATTATGG
GTCTTGGTACTATAAGTTCATCTCTCTCGTTGTTGACTTTTTAAGTCTCCAATAGTTTGT
TGTG

>G165 Amino Acid Sequence (conserved domain in AA coordinates: 7-62)
MVKKTLLGRRKVEIVKMTKESNLQVTFSKRKAGLFFKASEFCTLCDAKIAMIVFSPAGKVF
SFGHPNVDVLLDHFRCVVGHNNTNLDES YTKLHVQMLNKS YTEVKA EKEQKNQSR A
QNERENENAE EWWSKSPLELNLQSQSTCMIRVLKDLKKIVDEKAIQLIHQTNP NFYVGS S
NAAAPATVSGGNISTNQGFDDQNGMTTNPTQTLLFGFDIMNRTPGV*

>G1652 (77..1078)

AGCAAGTCCAAATCTCCCTCTCTCTCTCTATCTATCTCTATAGAAGATTTTTTAAC
TAAGAAGCTAGCGATCATGGCCACAGCGATGAACGTTTCTCTACCAAATGGTCTCCGA
ATTGGATATAGAAGAATATAGTATCATCCACCAATCCACATGAACCTACTCGTCGGAGA
TGTTCCACAGTCTCTCTCATCTCTTGATGATACCACCACTTGTTATAACCTTGATGCTTC
TTGTAATAAAAAGTTTGGGTAGAAGAAAGACCTTCAAAGATCCTCAAGACCACTCACATATC
ACCAAACCTTACATCCTTTTTCTCTCTCTAATCCTCCTCCTCCAAAGCACCAGCCCTCTTC
TAGGATTCTTTCTTTTGAAGACAGGTTTACATGTTATGAATCACAACCTCTCAAACCTT
AATATTTAGCCCCAAGGACGAAGAAATGGGATTACCAGAGCATAAGAAAGCCGAGCTGAT
AATAAGAGGGACAAAGAGAGCTCAATCCTTGACTCGAAGCCAATCAAATGCTCAAGATCA
CATACTGGCAGAGAGAAAACGGAGAGAGAAGCTTACTCAAAGATTTGTAGCTCTTCCGC
GCTAATTCCTGGCCTAAAGAAGATGGACAAGGCTTCTGTGTTGGGAGATGCAATAAAGCA
TATAAAGTACCTCCAAGAGAGTGTGAAGAGTATGAGGAACAAAAGAAGGAAAAGACAAT
GGAATCAGTGGTTCTTGTAAGAAGTCTAGTCTGGTTTATAGTGAATCATCAACCATC
ATCATCATCTTCTCAGATGGAAATCGCAATAGCTCGAGCTCAAATCTTCCAGAAATAGA
AGTTAGGGTTTTCAGGAAAAGATGTTCTTATTAAGATCCTATGCGAGAAGCAAAAGGGTAA
TGTGATCAAGATTATGGGGGAGATTGAAAAGCTTGGTTTGTCTATACCAACAGCAATGT
CTTGCCCTTTTGGACCACTTTTACATCTCTATTATCGCTCAGAAGAATAACAATTTGA
TATGAAAATCGAGGATGTTGTGAAGAACTTGAGTTTGGCTTATCAAAGCTCACTTAATT
GGTTTCACGTTACATACATATACACATTCATCATCGATTCTCCGATCGAAGAATCCAAA
ATCAGTTTTTCCATGAAAGTGGTTTTTTAGTTGTTAAGTTTGTGTTATGGAGATTCTTAA
GTCATTTAAAGATCCTTGTTCTTGTGTTTAAAGTGTGCTTAAAGATGCATATCATCAA
TGTTTAGTAATTATTTCTCTCCAGTTTCATTGGGACGGAATTTTTTTCGAGTTGTTGG
ATATATATTTCTGCGATGTAAAGCATTTCTGTTAGTTTAAATAACGTCGATATGTTTCT
TTGAAAA

>G1652 Amino Acid Sequence (domain in AA coordinates: 143-215)

MATAMNVFSTKWSSELDIEEYSIIHQFHMNSLVGDVPQSLSSLDLDTTCYNLDASCNKSL
VEERPSKILKTTTHISPNLHPPSSNPPPKHQPSRILSFECTGLHVMNHNSPNLIFSPK
DEEIGLPEHKKAELIIRGTRKRAQSLTRSQSNQDHLAERKRREKLTQRFVALSALIPGL
KKMDKASVLGDAIKHIKYLQESVKEYEEQKKEKTMESVVLVKSSSLVDENHQPSSSSSS
DGNRNSSSSNLPEIEVRVSGKDVLIKILCEKQKGNVIKIMGEIEKLGSLITNSNVLPGP
TFDISIIAQKNNNFMKIEDVVKNLSFGLSKLT*

>G1655 (132..755)

TTTCTAACTAGTCACATTGAGAGAGAGAGAGAGAGAAAAGAGAGACTCTCAGAATCTGAAG
AAGAAGAAGAGATTGTTGTTTTTGCCTTTTATCATCGGTTTCTTTGAATCTCTGGTTTTA
AATCGGATTTAATGGTGGAGTCTCTGTTCCCGAGCATCGAAAACACAGGTGAATCGTCTC
GAAGAAAGAAGCCGAGGATATCAGAGACGGCGGAGGCGGAGATAGAGGCACGACGTGTCA
ACGAAGAAAAGCTTTGAAGAGATGGAAGAACGAATCGTGTGCAACAGATCTACGCTTGTAAAGC
TCGTGCAAGCTTTACGCCGAGTTCGTGAGAGATCTCCACCACCAGCAACAACGAGACCG
ATAAATCTGCTCTCCGGCGCGGCGAGGGAGATACGTGATACGGCGGATCGAGTCTAGCTG
CGTCCGCTCGTGGTACGACTCGGTGGAGCAGAGCGATTTTAGCGAGTCGCGTCCGAGCGA
AGCTGAAGAAACATAGAAAAGCGAAAAAGTCAACGGGAAATTTGTAATCGAGAAAAGGTC
TCACGGAGACGAATCGGATTAAGTTACCGGCGGTTGAGAGAAAACCTGAAGATTCTTGGCC
GTTTGGTTCTCGGTTGCCGGAAGTCTCTGTACCGAATCTTTTAGATGAAGCGACCGATT
ACATCGCAGCGTTAGAGATGCAGGTTTCGAGCCATGGAGGCTCTCGCCGAACCTTTTAACCG
CAGCCGCACACCGGACGAGCTTGACCGGAACCTAACCGCGGCAGTTAGTTTGTGAGTTGT
TAATTAGCTTTTCTTTTACCTTTTACCCCTTTATTTTGGCTTCAAGTGTTTTTTTTTTC
TCGTGACGCGGATTTTAATTTATTAAATTCA

>G1655 Amino Acid Sequence (domain in AA coordinates: 134-192)

MVESLFPSIENTGESSRRKKPRISETAEAEIEARRVNEESLKRWKTNRVQQIYACKLVEA
LRRVRQRSSTTSNNETDKLVSGAAREIRDADRVLAAARGTTRWSRAILASRVRAKLKK
HRKAKKSTGNCKSRKGLTETNRIKLPAVERKLKILGRVPGCRKVSVPNLLDEATDYIAA
LEMQVRAMEALAEELLTAAAPRTTLTGT*

>G1671 (188..751)

TCCCACTATCCTTCGCAAGACCCTTCTCTATATAAGGAAGTTCATTTTCATTTGGAGAGG
ACACGCTGACAAGCTGACTCTAGCAGATCTGGTACCGTCGACCCTCTCTATATAATCTTC
TTCTACACACACACACACACGCAACCATATACGTACATGTGAAGTAGTGAGATCAATATC
GTTAGCAATGAATCTACCACCGGATTTAGGTTTTTTCCGACCGATGAAGAGCTCGTCGT
TCACTTCTCTCCACCGGAAAGCTTCCCTCTTGCCTTGTACCCCTGATGTATCCCCGACCT
TGATCTTTACCATTACGATCCTTGGGACCTTCCCGGAAAGCTTTGGGAGAAGGGAGGCA
ATGGTACTTCTATAGTAGAAAGACACAAGAGAGAGTGACAAGCAATGGGTATTGGGGATC
AATGGGAATGGACGAGCCAATCTACACAAGCTCCACACACAAGAAAGTGGGAATCAAAAA
GTATCTAACTTTCTATCTCGGAGATTCTCAGACTAATTGGATCATGCAAGAATATTCCCT
CCCGGATTCTCTTCTTCTATCTAGTCGATCTTCTAAGAGATCAAGCCGTGCTTCTAGTTC
TAGTCACAAACCCGATTATAGCAAGTGGGTGATATGCAGAGTGTATGAGCAAAATGTCAG
TGAGGAGGAAGACGATGATGGGACAGAACTCTCATGTTTGGATGAAGTGTTTTGTCTTT
AGATGATCTTGACGAAGTAAGCTTACCCTAATAAAGACAGAAGCACCCAAGAAGAGAAAA
AAAAAAAAAGGGTTAGTGGGCAATTATTTCTAAGCGACCGCTCTAGACAGGCCTAGTAC
CGGATCCTCTAGCTAGAGCTTTCTGTTCTGATCATCGGTTTCGACAACGTTTCGTCAAGT

>G1671 Amino Acid Sequence (domain in AA coordinates: TBD)

MNLPPGFRFFPTDEELVVHFLHRKASLLPCHPDVIPDLDLHYHDPWDLPGKALGEGRQWY
FYSRKTQERVTSNGYWGSMGMDPIYTSSTHKKVGIKKYLTFFYUGDSQTNWIMQEYSLPD
SSSSSSSRSSKRSSRASSSSHKPDYSKWVICRVYEQNCSEEDDDGTELSCLDEVFLSLDD
LDEVSLP*

>G1756 (71..1003)

ATATGTACTTGTACACCAACCCACCAAAAGAGATAAAAGAGGAAACAAAACTCGAAAAG
AGAGAGATATATGGGTGAGGTGGCTTATATGGACGAAGGAGACCTAGAAGCAATAGTCAG
AGGCTACTCCGGCTCCGGAGACGCGTTTTCCGGCGAAAGTTCCGGTACGTTTTTACCTTC
GTTTTGCTTACCGATCGAGAGCTCTAGTTTCTACGAACCGGAGATGGAGACAAGTGGCTT
AGATGAGCTCGGTGAACCTTTACAAACCTTTTACCCTTTCTCCACACAAACGATCCTCAC
AAGCTCGGTCTCTCTCCCTGAAGATTCAAACCTTTCCGAGATGACAAGAAACAACGATC
ACATGGTTGTCTTTTATCCAACGATCAAGAGCTGATCATATCCGAATTTCAGAATCCAA
ATCAAAGAAAAGCAAGAAGAATCAACAGAAGAGAGTTGTTGAGCAAGTGAAAGAAGAGAA

TCTGTTGTGCGGACGCATGGGCGTGGCGTAAATACGGGCAGAAACCCATCAAAGGATCTCC
ATACCCAAGGAGTTATTACAGATGCAGTAGCTCAAAGGGTGTGTTGGCAAGAAAACAAGT
CGAAAGAAATCCTCAAACCCGGAGAAATTCACCATAACATACACTAATGAGCACAATCA
TGAATACCAACCCGGAGAACTCATTAGCCGGTTCGACTCGAGCAAAAACCTCCCAACC
CAAACCAACCTTAAACCAAAAAATCCGAAAAAGAAGTTGTTTCTCCCTACAAGTAATCC
TATGATCCCATCCGCTGATGAATCTTCTGTTGCGGTTCAAGAAATGAGCGTTGCGGAAAC
GAGTACGCACCAAGCGGCTGGAGCAATCGAGGGCCGCGCTTGAGTAACGGTTTACCATC
GGATTTGATGTCCGGGAGCGAACTTTTCCAAGTTTTACCGGTGACTTCGATGAATATT
GAATAGCCAAGAGTTCTTCACTGAGTATTATGGAATTACTAGAGAGCATTAGGTGTATG
TATATATATAT

>G1756 Amino Acid Sequence (domain in AA coordinates: TBD)
MGEVAYMDEGDLEAIVRGYSGLDFAFSGESSGTFSPSFCLPMETSSFYEPMETSGLDDEL
GELYKPFYPFSTQTILTSSVSLPEDSKPFRDDKKQRSHGCLLSNGSRADHIRISESKSKK
SKKNQQKRVEQVKEENLLSDAWARKYGQKPIKGSPPYPRSYRCSSSKGLARKQVERN
PQNPEKFTITYTNEHNHELPTRRNSLAGSTRAKTSQPKPTLTKKSEKEVVSSPTSNPMIP
SADESSVAVQEMSVAETSTHQAAGAIERRLSNGLPSDLMMSGSGTFPSFTGDFDELLNSQ
EFFSGYLWNY*

>G1757 (250..1224)
ATCACCACCTCTATAACACTCTCATTCTCATCATATCATTCTTCAATCTATATAACCCAT
TCTTAATTATACTCAACACACATTATATTTTCTGATCATATCATTCTTTCAGTCCATCT
ATATAACCAATTCTTGATTTTATACTTAAACACACATTATACATCTTCTCATCATAGTT
TGTATCAATTTCCCTGAGTAACTACCTAAAGGAAAAAAAATCTATTTTGGGAATCAT
ATACTAAAAATGGAAGGAAGAGATATGTTAAGTTGGGAGCAAAAGACATTGCTAAGCGAG
CTTATCAATGGATTTGATGCGGCCAAAAAGCTTCAGGCACGACTTAGAGAAGCTCCGTCG
CCGTCGTCATCATTTTTCATCACCAGCGACGGCTGTTGCTGAGACTAACGAGATTCTGGTG
AAGCAGATAGTTTCTTCTACGAGAGATCTCTTCTTCTGCTAAACTGGTCATCCTCACC
AGCGTACAACTTATTCCGACGCCGGTTACTGTAGTCCCGGTGGCAAATCCCGGCAGTGTT
CCAGAACTCCGGCATCGATAAACGGAAGTCCGAGAAGTGAAGAGTTTCCGATGGAGGA
GGTTCTAGCGAGAGTCATCATCGCAAGATTACATTTTCAATTCAAAGAAAAGAAAGATG
TTACCAAAGTGGTCAGAAAAAGTGAGAATAAGCCCAGAGAGAGGCTTAGAAGGACCTCAA
GATGATGTCTTTAGCTGGAGAAAATATGGTCAAAAAGACATTTTAGGCGCCAAATCCCA
AGGAGTTATTACAGATGCACACATCGTAGCACACAAAAGTGTGGGCAACGAAACAAGTC
CAGAGATCAGACGGGGATGCTACGGTTTTTCGAAGTGACGTACAGAGGAACACACTTGT
TCGCAGGCGATCACAAGAACACCACCATTAGCCTCGCCGGAGAAGCGACAAGACACCAGA
GTCAAACCAGCCATTACCCAAAAGCCAAAGGATATTCTCGAGAGTCTTAAATCCAACTTA
ACCGTTTCGAACCGATGGGCTTGATGATGGTAAAGACGTTTCTCGTTCCCTGATACGCCG
CCGTTTTACAATTACGGAACCTATCAACGCGAGTTTCGCGCCAGTGGAGAGTTCTCCGATC
TTCGACGTTGTTGATGGTTCAATCCAACGGTCGAGATTGACACAACCTTCCCCGCGTTT
TTACACGAGTCGATTTATTATTAATTAATTAATTTGTAACAGAGAAATAGATAGTAAGT
AAGTAATGATCAGCGAGAGTTAAACATAAAAGTACTTAGAGTAATCTAACGATGCATAA
TAAGGAATGTTCAACAGGACTTGAACATGATTTCAATACTAAGAGAGATTTATCTAGCTA
CTGGTAGTAGCCGAGACTTCTTGTGTAGCTTCACTTNCCTTTTGTGCTT

>G1757 Amino Acid Sequence (domain in AA coordinates: 158-218)
MEGRDMLSWEQKTLSELINGFDAKKLQARLREAPSPSSSFSSPATAVAETNEILVKQI
VSSYERSLLLLNWSSSPSVQLIPTVTVVPVANPGSVPEPASPINGSRSEBFADEGGSS
ESHRQDYIFNSKKRMLPKWSEKVRISPERGLEGPQDDVFSWRKYQKDIILGAKFPRSY
YRCTRSTQNCWATKQVQRSDGATVFEVYRGTHTCQAIRTPPLASPEKQDTRVKP
AITQPKPDILESLSNLTVRTDGLDDGKDVFSFPDTPPFYNYGTINGEFGHVESPIFDV
VDWFNPTVEIDTTFPAFLHESIYY*

>G1782 (1..927)
ATGCAAGTGTTCAAAGGAAAGAAGATTCATCTTGGGGAACTCAATGCCTACAACAAAT
TCAAATATTCAAGGATCTGAATCTTTCAGCTTGACTAAGGATATGATAATGTCTACAACA
CAATTAACCCGATGAAACATTCCGGTTTGCAGCTGCAAAATCAAGATTCAACCTCATCA
CAATCTACTGAAGAAGAAATCAGGCGCGGTGAAGTTGCAAGCTTTGGAGAATATAAGCGT
TATGGATGCAGCATTTGTAATAACAATCTCTCAGGTTACATCGAAAACCTGGGAAAGCCT
ATTGAAAATTATACTAAGTCAATTACTACCTCGTCGATGGTGTCTCAAGACTCTGTGTTT
CCTGCTCCTACTTCTGGTCAAATATCTTGGTCTCTTCAATGTGCTGAAACGTCACATTC

AATGGTTTCTTGGCTCCTGAATATGCATCAACACCAACGGCGTGCCACATTTAGAGATG
ATGGGTTTGGTTTCTTCAAGAGTGCCATTGCCATCACATTCAAGAGAATGAACCAATA
TTTGTCAATGCGAAACAGTATCATCGGATTCTCCGTCGCAGGAAGCACCGTGCTAAACTC
GAAGCTCAGAACAACTCATCAAATGCCGTAAACCGTACCTTCATGAGTCTCGCCATCTT
CATGCTTTAAAGAGAGCTAGAGGCTCCGGTGGACGTTTCTCAATACAAAGAAGCTTCAA
GAATCATCAAACCTACTGTGTTCTTCTCAAATGGCAAATGGACAAAATTTCTCTATGAGC
CCTCACGGTGGTGGAGCGGAATCGGGTCTAGTTCGATCTCACCGAGCTCCAATTCAAAC
TGTATCAACATGTTCCAAAACCCGCAGTTCAGATTCTCAGGTTATCCGTC AACACACCCAT
GCCTCAGCTCTCATGTACGGGACTTGA
>G1782 Amino Acid Sequence (domain in AA coordinates: 166-238)
MQVFQRKEDSSWGNMPTTNSNIQSESPSLTKDMIMSTTQLPAMKHSGLQLQNQDSTSS
QSTEEESGGGEVASFGEYKRYGCSIVNNNLGYIENLGKPIENYTKSITTSSMVSQDSVF
PAPTSQIISWSLQCAETSHFNGLAPEYASTPTALPHLEMMGLVSSRVPLPHHIQENEP
FVNAKQYHAILRRRKHRAKLEAQNKLKCRKPYLHESRHLHALKRARGSGGRFLNTKKLQ
ESSNSLCSSQMANGQNFSMPHGGGSGIGSSSISPSSNSNCINMFQNPQFRFSGYPSTHH
ASALMSGT*
>G184 (327..1937)
TGAATTCTAGCCTTTTTGTAGGCGAATCATCTGGACCGGTAAGAGACTCTCTCATCGATA
ATAACCACATAATTTAATCAAACCTTTCTCTCTCTTTCTAAGATCTTTTGCTTTGCTCT
TTTCTTTTGTATCTTCTTATATATGGAGAAGCACAAAACGGTACTTACTATACGATAC
TGTACGGATCCATCAAACCTGGATTAATTATCAAACCGTACATTTTTTATCTTACCTGGCAA
GTTACATTCTAGGGTTTTGGAGAATCCAATCAACAACAAAGAAAATAATCATCGTTACA
ATAATCAGTATCACGCACAGACTTAGATGTTCCGGTTTCCAGTGAGTCTAGGCGGTTTAC
GTGACGAAGACCGTCACGATCAGATCACACCGTTGGATGACCATCGTGTGGTGGTTGATG
AGGTTGACTTCTTCTCAGAGAAGAGAGATAGGGTTTACGTGAGAACATCAACGACGACG
ACGACGAAGGCAATAAGGTTCTCATCAAAATGGAGGGTTCACGAGTTGAAGAAAACGATC
GTTCCAGAGATGTCAATATCGGTCTGAATCTTCTGACCGCGAATACGGGAACCGATGAGT
CAACGGTGGATGATGGACTATCAATGGATATGGAAGATAAAACGTGCAAAGATTGAGAACG
CACAAC'TACAAGAAGAGCTCAAGAAGATGAAAATAGAGAATCAAAGGCTAAGAGATATGT
TGAGCCAAGCGACGACCAACTTCAATGCCTTACAAATGCAACTTGTGTCCGTCATGAGGC
AACAAGAACAACGTAACCTTTCACAAGATCATCTCCTGGAGAGCAAAGCAGAAGGAAGGA
AACGGCAGGAACCTGCAAATCATGGTGC CAAGGCAGTTTATGGACCTTGGGCCGTCGTCGT
GAGCAGCAGAGCATGGAGCCGAAGTGTCTATCTGAAGAGAGGACAACGGTTCGTTTACGGTT
CTCCTCTCTCGCTTCTAGAAAGTTCCAATCCCGAGAGAACGGAAAGAGGTTGCTTGGAA
GAGAAGAAAGCTCAGAGGAATCAGAGTCTAACGCCTGGGGAACCCCTAACAAAGTCCCCA
AACATAATCCATCCTCTAGCAATAGCAATGGAAACAGAAACGGAAATGTTATTGATCAGT
CGGCCGCGAGAAGCCACCATGCGGAAAGCCCGTGTCTCAGTTTCGTGCCGATCTGAAGCTG
CCATGATAAGCGATGGATGTCAATGGAGAAAGTACGGACAAAAAATGGCTAAAGGAAACC
CGTGTCCGCGGGCTTATTATCGTTGCACAATGGCCGGTGGATGTCCAGTTTCGCAAGCAAG
TGCAGCGTTGCGCAGAAGACAGATCTATTCTCATAACCACCTACGAAGGAAACCACAACC
ATCCACTCCCACGACCGCTACGGCCATGGCCTCAACAACCACCGCAGCTGCAAGCATGC
TCCTCTCGGGCTCAATGTCTGAGTCAAGACGGTTTAAATGAACCCAACAACCTCCTAGCTC
GAGCTATCTTGCCTTGCTCCTCAAGCATGGCTACAATCTCAGCCTCCGCACCATTCCCAA
CCATCACATTGGACCTCACCAATTCACCAACCGGTAACAACCCCTAATATGACCCTAATA
ACCCGTTGATGCGAGTTTCGCTCAACGGCCCGGTTTCAACCCGGCAGTTTTGCTCAAGTGG
TTGGTCAAGCTATGTACAATAACCAACAACAGTCCAAGTTTTCTGGTTTACAGTTACCGG
CTCAGCCACTGCAGATCGCGGCCACTTCTCGGTGGCCGAGAGCGTTAGTGCTGCCAGTG
CAGCAATTGCGTCCGATCCAAACTTTGCGGCGGCTCTAGCGGCAGCGATCACGTCCATTA
TGAACGGTTCCAGTCTACAAAATAATAACACCAATAATAATAATGTGGCTACGAGCAACA
ATGACAGTAGGCAATAAGAGTTTTTCAATTTGATGGTTCGATTTTTTTTTTTGGGG
>G184 Amino Acid Sequence (domain in AA coordinates: 295-352)
MFRFPVSLGGSRDEDRHDQITPLDDHRVVVDVDFSEKRDVRSRENINDDDDDEGNKVL
KMEGSRVEENDRSRDVNI GLNLLTANTGSDESTVDDGLSMDMEDKRAKIENAQLQEELKK
MKIENQRLRDMLSQATTNFNALQMLVAVMRQEQRNSSQDHLLESKAEGKRQELQIMV
PRQFMDLGPSSGAAEHGAEBVSSEERTTVRSGPSPLLESSNPENGRKLLGREESSESE
SNAWGNPNKVPKHNPSSNSNGNRNGNVIDQSAEATMRKARVSVRARSEAMISDGCQW
RKYGQKMAKGNPCPRAYRYCTMAGGCPVRKQVQRCAEDRSILITTYEGNHNHPLPPAATA

MASTTTAAASMLLSGSMSSQDGLMNPNTNLLARAILPCSSSMATISASAPFPPTITLDTNS
PNGNNPNTMTNNPLMQFAQRPGFNPVLPQVVGQAMYNQSQSKFSGQLQPAQPLQIAAT
SSVAESVSAASAAIASDPNFAAALAAAITSIMNGSSHQNNNTNNNVATSNNDNRQ*

>G1845 (111..989)

AAGACATAATTTTCTCTGTTTTCTAGCTCTCTCTCTCAAATCTTCCATTGCTCTCTG
TTTTGGCAAATCGTGAAGTCCACGCTTTAAGGCATCAGTGAAGCAAAGATGGACTTTG
ACGAGGAGCTAAATCTTTGTATTACGAAAGGTAAAAATGTTGATCATTCTTTTGGAGGAG
AAGCTTCTTCCACGTCCTCAAGATCTATGAAGAAAATGAAGAGTCCTAGTCGTCCTAAAC
CCTATTTCCAATCCTCTCTCTCTCTCTTATTCTGTTAGAGGCTTTCCCTTTTTCTCTCGATC
CAACACTTCAGAATCAGCAACAACAACCTCGGATCATACGTTCCGGTACTTGAGCAACGAC
AAGACCCGACAATGCAAGGCCAGAAGCAAATGATCTCTTTAGTCCTCAACAACAACAAC
AGCAGCAGCAGTATATGGCCAGTACTGGAGTGACACATTGAATCTGAGTCCAAGAGGAA
GAATGATGATGATGATGAGCCAAGAAGCTGTTCAACCTTACATCGCAACGAAGCTGTACA
GAGGAGTGAGACAACGTCATGGGGAAAATGGGTGCGAGAGATCCGTAAGCCACGAAGCA
GGGCACGCTTTTGGCTTGGTACCTTTGATACAGCTGAAGAAGCTGCCATGGCCTACGACC
GCCAAGCCTTCAAATTACGAGGCCACAGCGCAACACTGAATTTCCGGAGCATTTTGTGA
ATAAGGAAAGCGAGCTGCATGATTCAAACCTCGTCCGATCAGAAAGAACTGAAACGCCAC
AGCCAAGCGAGGTTAACTTGGAGAGCAAGGAACTACCGGTGATTGATGTTGGGAGAGAGG
AAGGTATGGCTGAGGCATGGTACAATGCCATTACATCGGGATGGGGTCTTGAAAGTCTCTC
TTTGGGATGATTGGATAGTTCTCATCAGTTTTTCATCAGAAAGCTCATCTTCTTCTCTCTC
TCTCTTGTCTCTATGAGGCCTTTCTTTTGAAAAAGTTTATAAACCACATTGTGTTGTAGG
TTATAGTTTATAGGTTATGCTCATTGGCATTGGATGGAGGCAATTTTGTGATCTCCCAT
TCCACCACATATCAGTCATTATATGTGTCTACCTTTCTCTGTATTCTATCATTATCAT
TGTTTTTTATTATGTGTCTGTATGTGTTTCCCTATTGCTACATACATAGATGTCCTCTTTG
TTCAAAAAAAAAAAAAAAAAAAAAA

>G1845 Amino Acid Sequence (domain in AA coordinates: 140-207)

MDFDEELNLCITKGNVDHSFGGEASSTSPRSMKKMKSPSRPKPYFQSSSSPYSLFAFPF
SLDPTLQNOQQQLGSYPVLEQRQDPTMQGQKQMISFSPQQQQQQQYMAQYWSDTLNL
PRGRMMMMMSQEAQVQPIATKLYRGVRQRQWGWVAEIRKPRSRARLWLGTFDTEEAAM
AYDRQAFKLRGHSATLNFPEHFVNKESELHDSNSSDQKEPETPQPSEVNLESKELPVIDV
GREEGMAEAWYNAITSGWGPESPLWDDLDSSHQFSSESSSSSPLSCPMRPF*

>G1879 (3..917)

AAATGCCCTTAGAGGCTGTCTATACCCGCAAGATCCATTCCGATATCTCTCCAATTGCA
AAGATTTTATGTTCCACGACTTATACTCTCAAGAAGAGTTCGTAGCTCAAGATACGAAGA
ACAACATTGATAAGTTAGGGCATGAACAGAGCTTTGTGGAACAAGGTAAGGAGGACGATC
ATCAATGGCGAGACTATCATCAGTATCCTTTGTTGATCCCTTCGTTGGGAGAAGAGCTTG
GTCTTACCGCATTGATGTGGAGAGTCATCCTCCTCCACAGCACCGGAGGAAGAGGAGGA
GAACGAGAAAACCTGCAAGAACAAGGAAGAGATCGAGAACCAGAGAATGACTCACATCGCCG
TCGAGAGAAATCGCCGGAACAGATGAACGAGTATCTGGCTGTGCTCCGTTCTCTAATGC
CGTCGTCTGATGCTCAAAGAGGAGATCAAGCGTCGATAGTAGGAGGAGCTATAAACTACG
TGAAGGAGTTAGAGCATATTTACAATCTATGGAGCCGAAGAGAACTAGGACTCATGATC
CCAAAGGAGACAAGACTAGCACTAGCTCGTTAGTGGGTCCATTACAGATTTTTCAGCT
TCCCACAATATTCTACAAAGTCATCATCAGATGTACCGGAAAGCTCATCTTACCGGCGG
AGATAGAGGTTACGGTGGCAGAAAGCCATGCGAACATCAAGATAATGACGAAGAAGAAAC
CGAGGCAGCTTCTTAAGCTCATAACTTCTTTACAAAGCCTAAGGCTCACTCTTCTTCATC
TCAATGTCACCACTCTCCACAACCTCCATTCTCTACTCCATCAGCGTCAGGGTTGAAGAAG
GAAGCCAACCTGAATACCGTGGACGACATTGCAACAGCTTTGAATCAAACCATAAGGAGGA
TTCAAGAAGAGACATAATTAGCAAAATAGATTATAATTAAGTTGTTTTATTTTATTTTA
TTTTGAAATAACTGAAATCAGTTTTCTAATTTTTTTTTTTTTTCACTATTCCTCTAATCC
TCCCTATGTAAGTTGCATTTTGTCTCTTGTAATGAATCAATGGTCATAAAGATCTGAAC
AAAAAATTGAATAAAGAAATGGTT

>G1879 Amino Acid Sequence (domain in AA coordinates: 107-176)

MPLAEVVPQDPFGYLSNCKDFMFHDLYSQBEFVAQDTKNNIDKLGHEQSFEQKEDDH
QWRDYHQYPLLIPLSLGEELGLTAIDVESHPPPQHRKRRTNRNKNKEEIEHQRMTHIAV
ERNRRKQMNEYLAVALRSLMPSSYAQRGDQASIVGGAINVKELEHILQSMPEKRTTRTHDP
KGDKTSTSSLVGPFDFFSFPQYSTKSSSDVPESSSSPAEIEVTVAESHANIKIMTKKKP
RQLLKLITSLQSLRLTLHLNVTTLHNSILYSISVRVEEGSQLNTVDDIATALNQTIRRI

QEET*

>G1888 (1..729)

ATGAAGATTGGTGTGCTGTTTGTGATAAAGAAGAAGCTTCGGTGTTTTGTTGTGCGGAT
GAAGCAGCTCTTTGTAAATGGTTGCGATCGCCATGTTTCATTTTCGCCAATAAACTAGCCGGG
AAACATCTCCGGTTCTCTCTCACTTCTCCTACTTTCAAAGATGCTCCTCTTTGTGATATT
TGCGGGGAGAGGCGTGCATTATTATTTTGGCAAGAAGACAGAGCAATACTATGCAGAGAA
TGTGACATTCCAATACATCAAGCTAATGAGCACACTAAGAAACACAATAGATTCTCTCTT
ACCGGCGTTAAGATCTCTGCCTCCCCGTCAGCCTACCCAAGAGCCTCCAATTCCAACCTCT
GCTGCTGCATTTGGTTCGAGCCAAAACCCGACCAAAATCAGTATCGAGCGAGGTCCCAGC
TCGGCCTCCAATGAGGTATTTACGAGCTCTTCTTCGACGACCACGAGCAATTGCTATTAT
GGGATAGAAGAAAACCTACCATCACGTGAGCGATTCCGGGTCGGGATCGGGTTGTACAGGT
AGTATATCCGAGTATTTGATGGAGACATTACCGGGTTGGAGAGTGGAGGATTTGCTTGAA
CACCCTTCTTGTGTCCTCTATGAGGATAACATTATTACTAATAACAATAACAGTGAGTCT
TATAGGGTTTATGATGGTTCTTCACAATTCATCATCAAGGGTTTTGGGATCACAAACCC
TTCTCTTGA

>G1888 Amino Acid Sequence (domain in aa coordinates: 5-50)

MKIWCAVCDKEASVFCCADEAALCNGCDRHVHFANKLAGKHLRFSLTSPTFKDAPLCDI
CGERRALLFCQEDRAILCRECDIPIHQANEHTKKHNRFLTGVKISASPSAYPRASNSNS
AAAFGRAKTRPKSVSSEVPSSASNEVFTSSSSTTSSNCYYGIEENYHHVSDSGSGSGCTG
SISEYLMETLPGWRVEDLLEHPSVSYEDNIIITNNNNSES YRVYDSSQFHHQGFWDHKP
FS*

>G189 (34..987)

CCACAACCTCTCTCCTTGTAGAGAGAGAGATTTTATGGCGGTGGAGCTCATGACTCGGAAT
TACATCTCCGGCGTCGGAGCTGATAGCTTCGCCGTTCAAGAAGCAGCTGCTTCAGGACTC
AAAAGTATCGAAAAATTCATCGGTTTAAATGTCTCGTGATAGCTTTAACTCTGATCAGCCA
TCTTCTTCTTCCGCCCTCCGCCCTCCGCCCTCCGCCCGCCGAGATCTGAATCAGCTCGTAAC
ACAACGGCGGACGCGCGGTGTTTCAAAGTTTAAAAGAGTCATATCTCTCTTAGATCGAAT
CGAACCGGACACGCCCGGTTTAGACGTGCTCCGGTTTCAGTCTCCTCTCTCTCCGCCGCAAATGATC
CGAAAAGGTTCTGTTTCTTCATCGATGAAAACGATTGATTTCTCATCTCTCTCTCTGTA
ACAACGGAATCAGACAACCAGAAGAAGATTCATCATCATCAACGTCCCTCTGAAACGGCG
CCGTTTGGCTCTCAAACCTCAAAGCCTCTCCACGACGGTCTCGTCTTCTCAAATCAACA
AAGAGAAAATGTAACCTGAGAATCTTCTCACCGGAAAATGCGCTTCCGCTTCTTCTCTCC
GGTCTGTTGTCATTGCTCGAAGAAAAGATAAAACAGAGGAGAATAATTAGGGTTCCG
GCGATAAGTGCAAAATGTCCGATGTACCACCGGACGATTATTCATGGAGGAAATACGGA
CAAAAACCAATTAAAGGATCTCCACATCCAAGAGGATATTATAAGTGTAGTAGCGTAAGA
GGTGTGTCAGCACGTAAACATGTTGAGAGAGCAGCTGATGATTCTGTCATGTTGATTGTT
ACTTATGAAGGAGATCATAATCATTCTCTCTCCGCCGCTGATCTCGCCGAGCCGCCGTT
GCTGATCTTATTTTGGAAATCGTCTTGAAAAGAACAAATCTTTATTTAAGGCTTTTATAAT
ATAAATTTAGATCCTTACTTAGTGAAGTACTCAAACATGAATGAAATCAATGTAATCAA
AATCAAAAAGCTTTTGTCTAAAAA

>G189 Amino Acid Sequence (domain in AA coordinates: 240-297)

MAVELMTRNYISVGADSFVQEAASGLKSIENFIGLMSRDSFNSDQPSSSSASASASA
AADLESARNTTADA AVSKFKRVISLLDRTRTG HARFRRAPVHVISP VLLQEEP KTT PPFQS
PLPPPPQMIRKGSFSSSMKTIDFSSLSSVTESDNQKKIHHHQRPSETAPFASQTQSLST
TVSSFSKSTKRKCNSENLLTGKASASSSGRCHCSKKRKIKQRRRIIRVPAISAKMSDVPP
DDYSWRKYGQKPIKGSPPHGRYYKCSSVRGCPARKHVERAADDSSMLIVTYEGDHNHSL
AADLAGAAVADLILESS*

>G1939 (92..844)

AATCATTAGCTTCTTCTCTCTCTCTCACAGAGAGTAATCACAAGCCAAGTGAGA
AAAAGAAAACACTAAACCCAGATCGAAAACCATGTCTATTAACAACAACAACAACAA
CAACAATAACAAGATGGTCTTATGATCTCATCAAACGGAGCTTTAATCGAACAACAACC
ATCAGTCGTTGTGAAGAAAACACCGCGGAAAGATCGCATAGCAAAGTCGATGGAAGAGG
GAGAAGAATCCGTATGCCGATTATATGTGCTGCTCGTGTGTTTTCAGCTAACGAGAGAGCT
TGGTCATAAGTCAGATGGCCAAACAATTGAATGGTTACTTCGTCAAGCAGAGCCTTCTAT
TATAGCTGCAACAGGAACGTGTACAACCTCAGCGAGTTTCTCAACTGCTTCTGTCTCTAT
CCGTGGAGCCACCAATCTACTTCTTTAGATCATAAACCCACTTCTTTACTTGGTGGTAC

TTCTTGGCTTCACCATTTCTACCCCGGCGTTACGTCCACCCCGGCTTCTTCTGTCTCCCT
GCCACCACCACCAATGCTCCGCGTGAAGATGATATAGTGGAGCTTTTATGGCAAAGCGG
CCAAGTAGTTGGAACCAACCAACACATAGACAATCCTACGATCCTCCTCCCATTTCTCCG
CGGCAGCGGAAGTGGCAGAGGAGAAGAAAATGCTCCCTTTTCAACCTCCGCCTCACCT
GCATCAGCAAAATCTCTTCATTCAAGAAGGCGAAATGTATTCGTGGCTACACCATTTCTTA
CCGCCAAAATATTTCTGTCTCAGAACTTCTCAACTCCACTCCGGCTACTCACCCGCAAAG
TTCCATCTCTCTGGCACCACGTCACTATCGCCACGAGAAGGCGGAAAACTTTATGAA
CTTCTCGTGGCTAAGAGGGAACATATTTACCGCGGCTAGAGTTGATGAAGCTGGACCGTC
GTTTTCGGTGGTAAGAGAATCGATGCAGGTAGGCTCGAACACGACCCCCCTTCTTCTTC
TGCCACTGAATCATGTGTAATACCAGCTACAGAGGGCACCAGCGAGTGCAGTGTTCGGGAAC
TTTGGCAGCTCATGATCTTGGTCGGAAGGCGAAAGGCGGTGGCGGTTGAGGCGGCCGGAAC
ACCATCTTCAGGAGTGTGCAAGGCCGAAACAGAGCCGGTTTCAGATACAACAGCAACGGA
GTCGAAGCTAAAAGCGAGAGAAGAAACCCATGGAAGTGAAGAAGCTCGTGGTTCAACGTC
TAGAAAAGAGATCAGAACTGCAGAAATGCATAACCTCGCCGAAAGGAGAAGGAGAGAAAA
GATCAACGAGAAGATGAAGACTCTGCAACAACCTATTCTCGTGCACAAGGTTGAATC
TGATTCTGTTTCTACTCTGATCAGTCTACTAAAGTTTCAACGCTGGATGATGCTATCGAG
TACGTCAAATCGTTACAGAGCCAAATACAAGTATGCTCTTCAAAACAGAATGTGTTTAA
ACCAATGGTTCAACATGGAAAGAGTTTCATATGTATCTAGTTTTGTTGAGATGATGTCGAC
GGGACAGGGTATGATGTCGCCAATGATGAATGCCGGAATACGCAACAGTTTCATGCCCA
TATGGCCATGGATATGAACCGACCTCCTCCATTTCATACCTTTCCCGGCACATCTTTTCC
TATGCCGGCTCAAATGGCAGGTGTAGGTCCATCATATCCAGCACCGCGCTACCTTTTCC
CAACATTCAGACCTTTGACCCATCCAGAGTCCGTTTACCAAGCCCGCAGCTTAACCCGGT
GTGCAACAGCCTCAGTTTCCGGCTTACATGAATCCCTATAGCCAGTTTGTCTGGTCCCA
CCAGTTGCAACAACCTCCTCCTCCTCCATTTCAGGCTCAAACAACATCACAACCTGAGTTC
CGGGCAGGCAAGTAGTAGCAAGGAACCTGAGGATCAGGAGAACCAACCAACAGCTTAGTT
AAAGTGTGGAGCTGAAACGGATCAGTTCTTCAAGCAAATTACAACCTTGAAGATAAACCA
GAGTTGTAACATGTAGATTTTGTCTGTTAAGTTTAAATGTAAGTACTTTTATGTTAATGGG
AAAGATACTGACAGGTTGCAAGGTGGTCAGTATTTGTGCATCAGCTTAAGATTCTCGA
TGTGGCCAGTATCTCCCTTTTCTAGCATGTGAGGTCCCTACTCTCTGGTTCTACGGAGAC
CAAATGTTGACTGATTAAACACACAATGACTTACCAAAAGTACACGCGGCCCATCTCG
TCTTTATGTTCCAAGTGCAGCTGTTTGTATTTGTAAGCATTTTTCTTATAATAATAAA
ACAGCTCTATCTTCGTTAAAAAAA

>G1943 Amino Acid Sequence (domain in AA coordinates: 335-406)

MGEDDIVELLWKSQGVVRSQTQRPSSNTPPSLPPPIILRGSGSGNGEENAPLPLPQSP
PLHHQNLFILEDEMSSWLHSHPGVSTPASSVSLPPPNAPREDDIVELLWQSGQVVG
NQTHRSYDPPPIILRGSGSGRGEENAPLSQPPPHLHQNLFIEGEMYSWLHHSYRQNYF
CSELLNSTPATHPOSSISLAPRQTIATRRANFNMNFWLGRNIFTGGRVDEAGPSFSVVR
ESMQVGSNTTPPSSSATESCVIPATEGTASRVSGTLAAHDLGRKGKAVAVEAAGTPSSGV
CKAETEPVQIQPATESKLKARETHGTEEARGSTSRKRSRTAEMHNLAERRRREKINEKM
KTLQQLIPRKNKVESDSVSTLISLLKFQRWMLSSSTSNRYRAKYKYLQNRMCFKPMVQH
GKSSYVSSFVEMMSTGQGMSPMMNAGNTQQFMPHMAMDMNRPPPIPFPGTSFPMQAQM
AGVGPSYPAPRYFPNIQTFDPSRVRLPSPQPNPVSNPQFPAYMNPYSQFAGPHQLQP
PPPPFQGGTTSQLSSGQASSSKEPEDQENQPTA*

>G21 (79..966)

TGTGGAGGAATATTAATACAGCCCACTTCACATCTATTTTGTGCAACCATCTCTCTAA
GCTTCTTCTCTCATAACAATGGCAAGACAAATCAACATAGAGAGTAGTGTCTCAAGTT
ACCTTTATCTCCTCCGCATCCCGCCGTATCTTCTCCTCCTCCATCACCGCTTCCGCC
TCATTGTCCTCTTCACCTACTACATCTTCTCTTCTTCGTATCAACAAATCTAACTTC
ATTGAGGAAGACAACCTCTAAAAGAAAAGCATCTCGAAGATCATTGTATCGTTAGTCTCC
GTTGAAGACGATGATGATCAAAACGGTGGAGGTGGGAAACGGCGAAAGACCAACGGTGA
GATAAACATCCGACGTATAGAGGAGTGAGGATGAGGAGTTGGGAAAATGGGTGTTCGGAG
ATTAGAGAGCCGAGAAAGAAATCAAGAATCTGGCTCGGGACTTATCCAACGGCTGAGATG
GCAGCTCGAGCTCATGACGTAGCGGCTTTAGCCATTAAAGGTACAACGGCTTACCTCAAT
TTTCCCAAGTTAGCCGGCGAGCTTCTCTCGTCCAGTCACAAATCTCCTAAAGACATTCAA
GCCGCCGCTCTTTAGCGGCCGTTAACTGGCAAGATTCCGTCAACGATGTGAGTAATTCT
GAAGTGGCTGAAATAGTTGAAGCCGAGCCGAGTCGAGCCGTGGTGGCTCAGTTGTTTTCT
TCGGACACAAGCACGACGACGACTCAGAGTCAAGAGTATTCGGAAGCTTCGTGTGCT

TCGACTTCGGCGTGACGGACAAAGACAGTGAGGAAGAGAAGCTGTTTGATTTGCCGGAT
TTGTTTACCGATGAGAATGAGATGATGATACGAAACGATGCGTTTTGCTACTACTCGTCC
ACGTGGCAGCTTTGTGGAGCCGATGCTGGGTTTCGGCTTGAAGAGCCGTTTTTCTATCT
GAATGACTAAAGTACCCCTCTCGAGAGAGCTCTCACTAACACT

>G21 Amino Acid Sequence (domain in AA coordinates: 97-164)
MARQINIESSVSQVTFISSAIPVSSSSSITASASLSSSPPTSSSSSSSTNSNFIEEDNS
KRKASRRSLSSLVSVEDDDDDQNGGGGKRRKTNGGDKHPTYRGVVRMRSWKVSEIREPRK
KSRIWLGTYPTAEMAARAHDAALAIKGTAYLNFPKLAGELPRPVNTNSPKDIQAAASLA
AVNWQDSVNDVSNSEVAEIVEAEPSRAVVAQLFSSDTSTTTTTQSQEYSEASCSTACT
DKDSEEEKLFDPDLFTDENEMIRNDAFCYSSSTWQLCGADAGFRLEEPFFLSE*

>G2132 (42..1031)

ATTCTGTTACTTAGTACCGGAGTTTAGTCGGAGAGAGAACAATGATCAGTTTCAGAGAAG
AGAACATCGATCTCAACTTGATTAAAAACAATTAGTGTAATCTGTAATGATCCAGACGCCA
CCGATTCCCTCTAGCGACGATGAATCTATCTCCGCAATAATCCTCGCCGTCAGATCAAAC
CAAACCACCGAAACGTTACGTCTCAAAGATCTGTGTCCCGACGCTGATCAAAGGTATG
AGAACGTTTCGAATTCTACAGGAATAAAGCAGCCGAAACCGGAAAACGTCGTCGGGTT
TCAAAGGCGTACGACGGAGGCCGTGGGGGAAATTTCGGCGGAGATAAGAAATCCGTTTG
AGAAGAAGAGAAAGTGGCTTGGAAACGTTTCTACTGAAGAAGAAGCAGCAGAAGCTTACC
AAAAGAGTAAAGAGAGTTTGATGAACGATTGGGTTTAGTTAAACAGGAAAAAGACCTAG
TAGATTTGACCAAGCCGTGCGGTGTACGTAAACCAGAAGAGAAGGAAGTTACTGAGAAGT
CGAATTGCAAAAAGGTAAATAAGAGAATTGTTACTGATCAGAAGCCATTGGTTGTGGTT
ATAACGCTGATCAATGAAGAAGAGGGAGTGATTAGTAAAATGTTGGAAGATCCGTTGATGA
CATCGTCAATTGCTGATATTTTTGGTGATTCCGCTGTTGAAGCAAATGATATTTGGGTGG
ATTACAATTCAGTGGAATTTATTTCCATTGTAGATGATTTCAAGTTTGATTTGTGGAGA
ATGATAGAGTAGGAAAGGAGAAAACATTTGGATTTAAGATTGGGGATCACACTAAAGTTA
ATCAACATGCCAAAATCGTATCGACCAATGGGGACTTATTCGTCGATGATTTACTTGATT
TTGATCCGTTGATAGATGATTTTAAAGTTAGAAGATTTTCCTATGGATGATCTTGGATTAT
TAGGAGATCCAGAGGATGATGATTTTAGTTGGTTTAAAGTTACTACTGATTGGATCGATA
AGTTTTTATGAATACTTTTCTTGACACGGCCAACGGTATTAGTAC

>G2132 Amino Acid Sequence (domain in AA coordinates: TBD)
MISFREENIDLNLIKTISVICNDPDATDSSSDDESISGNNPRRQIKPKPKRYVSKICVP
TLIKRYENVSNSTGNKAAGNRKTS SGFKGVRRRPWGKFAABIRNPFKKRKLWLTGTFPTEE
EABEAYQKSKREFDERLGLVKQEKDLVDLTKPCGVKPEEKEVTEKSNCKVKNRIVTDQ
KPF GCGYNADHEEEGVISKMLEDPMTSSADI FGDSAVEANDIWDVNSVEFISIVDDF
KFDFVENDRVGKEKTFGFKIGDHTKVNQHAKIVSTNGDLFVDDLDFDPLIDDFKLEDFP
MDDLGLLGDPEDDDFSWFNGTTDWIDKFL*

>G2145 (1..777)

ATGGACGTTTTTGTGTGATGGTGAATTGGAGTCTCTCTTGGGGATGTTCAACTTTGATCAA
TGTTTCATCATCTAAAGAGGAGAGACCGCGAGACGAGTTGCTTGGCCTCTCTAGCCTTTAC
AATGGTCATCTTCATCAACATCAACACCATAACAATGTCTTATCTTCTGATCATCATGCT
TTCTTGCTCCCTGATATGTTCCCATTTGGTGCAATGCCGGGAGGAAATCTTCCGCCATG
CTTGATTCTTGGGATCAAAGTCATCACCTCCAAGAAACGCTCTTCTTAAAGAGGAAACTA
CTTGACGTGGAGAATCTATGCAAACTAACTCTAAGTGTGACGTACAAAGACAAGAGCTT
GCGAAATCCAAGAAAAAACAGAGGGTAAGCTCGGAAAGCAATACAGTTGACGAGAGCAAC
ACTAATTGGGTAGATGGTCAGAGTTTAAGCAACAGTTTCAGATGATGAGAAAGCTTCGGTC
ACAAGTGTTAAAGGCAAACTAGAGCCACCAAGGGACAGCCACTGATCCTCAAAGCCTT
TATGCTCGGAAACGAAGAGAGAAGATTAACGAAAGGCTCAAGACACTACAAAACCTTGTC
CCAAACGGGACAAAAGTCGATATAAGCAGATGCTTGAAGAAGCGGTCCATTACGTGAAG
TTCTTGACAGCTTCAGATTAAGTTGTTGAGCTCGGATGATCTATGGATGTACGCACCATTG
GCTTACAACGGCCTTGACATGGGGTTCCATCACAACCTTTTGTCTCGGCTTATGTGA

>G2145 Amino Acid Sequence (domain in AA coordinates: 166-243)
MDLVFDGELESLLGMFNFQDQSSSKEERPRDELLGLSSLYNGLHQLHQQHHNNVLSDDHHA
FLLPDMFPFGAMPFGNLPAMLDSWDQSHHLQETSSSLKRKLLDVENLCKTNSNCDVTRQEL
AKSKKKQRVSSSNTVDESNTNWVDGQSLSNSDDEKASVTSVKGKTRATKGTATDPQSL
YARKRREKINERLKTLLQNLVPNGTKVDISTMLEEAVHYVKFLQLQIKLLSSDDLWMYAPL
AYNGLDMGFHNNLLSRLM*

>G23 (22..732)

TATCAAACGAGAGTACAAAAGATGACGTCCTCAACAGCTCTGCATCACCAACATCATCG
TCATCAGACCAATCTGATGCAACTACTACAACAAGCACCCTTGTCTGAAGAAGAAGCT
CCACCCAGAAACAACAACAAGAAAGAGAAGGAGAGATTCTTCTCTGCTTCTTCATCT
TCTTCAATGCAACATCTGTTTACAGAGGTGTGCGGATGAGAAGTTGGGGCAAATGGGTC
TCCGAGATCCGACAACCTCGTAAGAAAACCTGATTTTGGCTCGGCACCTTTTGTACCGCT
GATATGGCTGCTCGTGCTCACGACGTGCTGCTCTACCATCAAAGGCTCCTCCGCCGTC
TTAAATTTCCCTGAGCTTGCTTCTCTCTTCCCTCGTCCGGCGTCATCATCGCCGCATGAT
ATCCAGACAGCCGCCGAGAAGCCGCCGCCATGGTGGTGAAGAAAACTGTTAGAGAAG
GATGAGGCTCCGGAGGCCACCTTCGTCGGAATCTTCTTACGTGGCGGCGGAGTCAGAG
GATGAGGAGAGGTTGGAGAAAATTGTGGAGCTGCCTAACATTGAAGAAGGAAGTTATGAC
GAGAGTGTGACATCACGTGCTGATCTGGCTTATTCTGAGCCGTTTCGATTGTTGGGTGTAT
CCTCCGGTTATGGATTTTTATGAAGAAATATCGGAGTTAATTTCTGGAATTGTGGAGC
TTTATCTTCTTTTCTTTTGTACGGATACGTAGCATATGCTACTATACTAGGGCTAGTG
TACCAAATATTGTAATAATACTTATTAATATTTATGTAAATGTGTAATATATATAACAT
ACAATTATTGTAAGTTTGGAAATGGAACTATCGTTACGCAATGTTCTGTAAAAAAA
AAAAAAAAA

>G23 Amino Acid Sequence (domain in AA coordinates: 61-117)
MTSLNSSASPTSSSSDQSDATTTSTHLSEEEAPPRNNTRKRRRDSASSSSSMQHPV
YRGVMRWSWKWVSEIRQPRKTRIWLGTFVTADMAARAHDAALTIKSSAVLNFPELA
SLFPRPASSPHDIQTAAAEAAAMVVEEKLLEKDEAPEAPSSSESYVAAESEDEERLEK
IVELPNIEGSDYDESVTSRADLAYSEPFDCWVYPPVMDFYEEISEFNFVELWSFNH*

>G2313 (104..724)
CGTCGACACAATCGCTCTTCCGTAACATATTCACAAAACGATCTTCTTGTTCCTTGAAT
TTTTAGCCATCTCTTTTTTTTTTTTCTCATTTTCTCGGATACTATGGCTTCGAGTCCACG
CTGGACGGAGGACGACAACAGGCGTTTAAAGTCAGCTCTGTGCAATTCCTCCGGATAA
CAAGCGTTTGGTGAATGTGCGCCAGCATCTGCCGAAACCTTTGGAGGAGGTGAAGTACTA
CTACGAAAAGTTGGTCAACGATGTTTATCTGCCGAAACCTTTAGAGAATGTCAACCCAGCA
TGTTAATCAGATGCCCGAGTACGTACCACTGGCGGAATCGAGTCAGTCCAAACGCAGGAA
GAAGGATACGCCAAATCCTTGGACAGAAGAGGAACACAGATTGTTTCTGCAAGGATTGAA
AAAGTATGGGGAAGGAGCTTCGACGTTGACATCAACGAATTTGTGAAGACAAAGACTCC
ACGGCAAGTGTAAGCCATGCACAGTATTACAAAAGGCAAAAATCGGACAAATAAGAAGGA
GAAACGCCGGAGTATTTTTGACATAACTTTGGAGTCTACCGAGGGCAATCCAGATTCTGG
AAATCAGAACCTCCGGATGATGATGATCCGTCCCAAGGTCAAGGCACCTTGTCTTGGAGT
TTAGATGTTGGAAGATAGAAGAATGGTGTGAAAGC

>G2313 Amino Acid Sequence (domain in AA coordinates: TBD)
MASSPRWTEDDNRRFKSALSQFPDPNKRVLNVAQHLPKPLEEVKYYEKLVDVYLPKPL
ENVTLQHLQKPMEMEEMKMYEKMANDVNQMPEYVPLAESSQSKRRKKTDPNPWTREEHRL
FLQGLKKYGEASTLTSTNFVKTTPRQVSSHAQYKRQKSDNKKEKRRSIFDITLESTE
GNPDSGNQNPDDDDPSQGGTCLGV*

>G2344 (1..573)
ATGACTTCTTCAATCCATGAGCTTCTGATAACATTGGAAGTCATGAGAAGCAAGAACAG
AGAGATTCTCATTTCCAACCACCAATCCCTTCTGCAAGAAATTATGAATCAATTGTTACA
AGTTTAGTCTACTCAGACCCGGGACTACAAATTCCATGGCACCTGGACAATATCCATAT
CCAGATCCTTACTACAGAAGCATATTTGCACCGCCTCCACAACCGTATACCGGGGTACAT
CTACAGTTGATGGGAGTGCAGCAACAAGGCGTTCCTTTACCATCTGATGCAGTCGAGGAA
CCTGTTTTTGTAAACGCAAGCAATACCACGGTATACTAAGGCGCAGACAATCAAGAGCA
AGACTTGAGTCTCAGAATAAAGTCATCAAGTCACGTAAGCCGTATTTGCATGAATCTCGG
CATTGTCATGCGATAAGACGACCAAGAGGATGTGGCGGGCGGTTTCTAAATGCCAAGAAG
GAGGATGAGCATCACGAAGACAGTAGTCATGAAGAAAAATCCAACCTTAGCGCTGGTAA
TCCGCCATGGCTGCTTCTAGTGGTACATCTTGA

>G2344 Amino Acid Sequence (domain in AA coordinates: TBD)
MTSSIHELSDNIGSHEKQEQRDSHFQPPIPSARNYESIVTSLVSDPGTTNSMAPGQYPY
PDPYRYSIFAPPPQPYGTGVHLQLMGVQQQGVPLPSDAVEEPVFVNAKQYHGILRRRQSRA
RLESQNKVIKSRKPYLHESRHLHAIRPRGCGGRFLNAKKEDEHHEDSSHEEKS NLSAGK
SAMAASSGTS*

>G2430 (69..1907)

AACTTCAACATACACATAATCTCTCACTTAAAAATATCTCTCTCTCTCTCTACAAAAT
CAATTTCCAATGTTGGTGGGAAAGATAAGTGGATATGAAGATAATACTCGCTCTTTGGAGC
GAGAAACATCTGAAATCACTTCTCTCTCAGCCAAATTTCCGGGGAATACTAATGTCCCTTG
TTGTTGACACCAATTTACCACCTCTACTCAACATGAAACAAATCATGAAACAATACGCTT
ATCAAGTGTCTATTGAGACAGATGCAGAAAAAGCTCTTGCGTTTTTGACAAGCTGCAAAC
ATGAAATCAATATTGTGATTTGGGATTTTCATATGCCTGGAATTGATGGACTTCAAGCTC
TCAAGAGCATTACTTCAAAGTTGGATTTACCTGTAGTGATTATGTCTGATGATAATCAAA
CGGAATCTGTGATGAAAGCAACATTTTACGGTGCTTGTGACTATGTTGTGAAACCGGTTA
AAGAAGAGGTAATGGCCAATATATGGCAACACATTTGTACGGAAGAGGCTGATCTTTAAAC
CGGATGTTGTCTCCACCGGTTCAATCAGATCCGGCTCGCTCTGACCGTTTAGACCAAGTCA
AAGCTGATTTCAAGATCGTAGAAGATGAACCAATAATCAATGAGACACCGCTGATCACAT
GGACCGAAGAAATTAACCGGTTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAG
TGAATGGCTATTCCCCAATCATGAACCAAGATAACATGTTCAACAAAGCACCACCTAAAC
CGCGAATGACGTGGACAGAAGTTATTCAACCGGTTCAATCAAATCTGGTTCAAACAAAAG
AGTTCCGGCCAACTCAATGACTATTCCCAAATCATGAACCAAGATAGCATGTACAACAAAAG
CAGCAACCAACCAACAATTGACGTGGACCGAAGAAATTAACCGGTTCAATCAGGTCTGG
TTCAAGCCAACGAGTTTCAAGTGAATGGATATTCCCAAAGCATGAACCAAGATAGCA
TGTTCAACAAATCAGCAACCAACCCGCGATTGACATGGAACGAATTACTTCAACCGGTTT
AATCAGATCTGGTTCAATCCAATGAGTTTAGCCAATTCAAGTCAATCTCTCAAATCATGA
ACGAAGATAACATGTTTCAACAAAGCAGCAAGAAACCGCGGATGACATGGAGTGAAGTAT
TTCAACCGGTTCAATCACATCTGGTTCCGACTGACGGTTTAGACCGAGACCACTTTGATT
CCATAACCATAAACCGAGGTAACGGCATAACAAACATGAAAAGAAACAAGGAAAAAAAC
CACGGAAGCCGCGGATGACGTGGACCGAAGAGCTTACCAAAAAATTTCTGGAAGCCATCG
AAATAATTGGTGGTATCGAAAAAGCTAACCCAAAGGTACTTGTGCAATGCTTGCAAGAAA
TGAGGATAGAAGGAATTACTAGAAGCAATGTGGCAAGTCATCTTCAGAAACACCGTATCA
ATCTTGAAGAAAACCAATTCCTCAACAAACACAAGGGAATGGTTGGGCCACTGCGTATG
GTACACTAGCTCCCTCTCTCCAAGGTTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAG
TTATGAATGGTCCAGCCACTTTGAACCAAAATCCAGCAGAAATCAATATCAAATGGTTTCT
TGACAATGAACAACAACAGATCATAACCAATCCTCCGCCCTCTTTGCCCTATTTGGACC
ATCATCACCAACAGCAACATCAGTCTTCTCTCAATTTAATTACCTGATGAACAATGAAG
AACTTCTTCAAGCTCTGGCCTCTCTGCGACAGATCTTGAACCTCACTTATCCAAGTTTAC
CATATGATCCACAAGAGTATCTAATCAATGGCTACAATTATAATTAGTCATATAGCCCTT
CTCTTTACTTAAGGCAGTCTATGTATGACAAATAATATGCGACTTCCCTTGTGAGTCACA
ATATTGTTTCATTATTC

>G2430 Amino Acid Sequence (domain in AA coordinates:425-478)

MLVGKISGYEDNTRSLERETSEITSLLSQFPNGNTNVLVVDNFTTLLNMKQIMKQYAYQV
SIETDAEKALAFLTSCKEHINIVIWDFHMPGIDGLQALKSITSKLDLPVVIMSDDNQTES
VMKATFYGACDYVVKPVKEVMANIWQHIVRKRLIFKPDVAPPVQSDPARSDRLDQVKAD
FKIVEDEPIINETPLITWTEEIQPVQSDLVQANKFDQVNGYSPIMNQDNMFNKAPPKPRM
TWTEVIQPVQSNLVQTKBFGQLNDYSQIMNQDSMYNKAATKPQLTWTEEIQPVQSGLVQA
NEFSKVNQYSQSMNQDSMFNKSATNPRLTWNELLQPVQSDLVQSNEFSQFSDYSQIMNED
NMFNKAACKPRMTWSEVFQPVQSHLVPTDGLDRDHFDSTINGGNGIQNMEKKQKKPRK
PRMTWTEELHQKFLAIBIIGGIEKANPKVLVECLQEMRIEGITRSNVASHLQKHRINLE
ENQIPQQTQGNQWATAYGTLAPSLQGSNDVNTTIPSYLMNGPATLNQIQNQYQNGFLTM
NNNQIITNPPPLPYLDHHHQQQHQQSSPQFNYLMNNEELLQASGLSATDLELTYPPLPYD
PQEYLINGYNYN*

>G2517 (66..899)

TCCTCACTCTCTCTCTTTTCTCTAACCATAAAATCTCTTTGATCTCTTTCTCTGTGTTT
TGATAATGGAATGTTGGTGTGGGATGCCGTTTTACGATTTAGGGCAAACAAGGGTTT
ACCCACTCTTGTCTGATTTCCACGATTTATCGGCGGAGAGGTATCCGGTAGGGTTCATGG
ATTTACTGGGTGTTTCATCGTCATACACCCACCCATACGCCGTTGATGCATTTTCCGACCA
CACCTAATCGTCTCGAGCGAAGCTGTGAATGGAGATGACGAAGAAGAAGATGGAG
AAGAACAGCAGCATAAGACAAAGAAGCGGTTAAATTCACATAAATGAGTAGAAAGCAGA
CGAAGAAGAAGTGCCAAAAGTGTCTATTCATCACGAGGAGTGAGGTTCTTCATCTAGATG
ATGGTTATAAGTGGAGAAAATACGGTCAAAAACCTGTCAAAGACAGCCCTTTTCCAAGAA
ATTATTACCGTTGCACAACAACCTTGGTGTGACGTGAAGAAGAGAGTAGAGAGATCATTCA

GTGATCCAAGCAGTGTAATCACCACCTTACGAAGGTCAACATACTCATCCTCGTCCACTAC
TCATCATGCCCAAAGAAGGCAGCTCTCCATCCAATGGCTCAGCTTCTAGGGCCACATTG
GCCCTCCCTACACTCCCTCCTCAGCTTTTAGATTACAACAACCAACAACAAGCGCCGT
CTTCTTTTGAACCGAGTACATTAACAGGCAAGAAAAAGGAATTAATCATGATGATGATG
ACGATCATGTTGTGAAGAAGAGTCAAACTCGGGATCTGCTGGATGGAGCTGGTTTAGTCA
AAGATCATGGCCTTCTTCAGGATGTTGTTCCCTCTCATATCATTAAAGGAAGAGTATTAGT
TAATCGCATAATTATGTAGCTAGCTAGCTAG

>G2517 Amino Acid Sequence (domain in AA coordinates: TBD)

MENVGVGMPFYDLGQTRVYPLLSDFHDLAERYPVGFMDLLGVHRHTPTHTPLMHFPTTP
NSSSSEAVNGDDEEEDGEEQHKTKRKFKTKMSRKQTKKKVPKVSFIFRSEVLHLLDDG
YKWRKYGQKPKVDSPPFRNYRCTTTWCDVKRVERSFDPSVITTYEGQHTHPRPLLI
MPKEGSSPSNGSASRAHIGLPTLPPQLLDYNNQQQAPSSFGTEYINRQEKGINHDDDDDD
HVVKSRTRDLLDGAGLVKDHGLLDQVVPShiIKEY*

>G2521 (103..768)

ATTCTCCACAATTTTATAACTTTCTTCCGCTCAACTTCAGATAAAATTCGGATTCTGTAGC
TCTTTTCAATACGACTGCGGAGATCAGAGCCAATTATTTGGTTATGGCGTCTCTGATCTCA
GATATTGAACCGCCGACGAGTACTACTTCAGATCTCGTTCCGAGAAAAGAGATCCTCT
GCTTCATCCGCGCATCGTCTCGTTCAAGCGCATCTTCCGTCTCCGGTGAGATTCACGCG
CGATGGCGATCGGAGAAGCAACAACGGATCTACTCAGCCAAACTGTTCCAAGCGCTCCAA
CAAGTCCGCTCAACTCTTCCGCTCAACATCATCATCTCCAACGGCTCAGAAACGAGGA
AAGGCCGTCCGTGAAGCCGCCGATCGAGCTCTTGCCGTCTCCGCTCGGGGAAGAACTC
TGGAGCAGAGCGATCTTAGCTAATCGGATCAAACTGAAATTTCTGTAACAGAGACGTCCT
CGAGCTACGATGGCGATTCCGGCCATGACTACGGTGGTTAGTAGCAGCAGCAACAGATCG
AGAAAACGAGAGTGTTCGGTGTGAGATTGAATAAGAAGAGTATACCGATGTTAACCGG
AAAGTACGTGTTCTAGGCCGGTTAGTTCCCGGTTGCGGTAAACAATCCGTACCGGTGATT
CTAGAAGAAGCAACTGATTATATTCAGGCTCTGGAGATGCAAGTGAGAGCCATGAACTCT
TTAGTTTCAGCTTCTCTCTCTACGGCTCAGCTCCTCCACCGATTGATGAGGTTAAAT
CGTCTTTTAAATTCTACCATCTCTCGATCTTTCACAGCTTATGTGTATATAGAAGATTCTG
GTTTGATTATAATCTGTAACACTCTTCCCAACCGCTGATTCTTCTCTGCTACAAGTAAA
AGTAAATTTGAACCGAGTCTTCCCATTTTTACGATCCTCAAGTCTAAATTAAGTATATG
ATTGATTAATAAAGTCTTTACCATTAGGGTTC

>G2521 Amino Acid Sequence (domain in AA coordinates: 145-213)

MASLISDIEPPTSTTSDLVRRKRSSASSASSRSSASSVSGEIHARWRSEKQORIYSAK
LFQALQQVRLNSSASTSSSPTAQKRGKAVREAADRALAVSARGRTLWSRAILANRIKLF
RKQRRPRATMAIPAMTTTVSSSSNRSRKRRVSVLRNLNKKSIPTVNRKVRVLGRLVPGCGK
QSVPVILEEATDYIQALEMQVRAMNSLVQLLSYGSAPPPPI*

>G258 (60..983)

AGTGACCACCCTGCTGGTTAATCAACACCAAGAGACCTTGTAATATATAAGTTAGGAAGA
TGAGAGAGAAGTGGGAAATGAAAAGAGATGAAATGGGACATCGATGTTGTGGAAAACACA
AAGTGAAGAGAGGTCTTTGGTCTCCAGAGGAAGACGAGAAGCTTCTCGTTATATCACCA
CTCATGGTCATCCTAGTTGGAGTTCCGTTCCAAAGCTTGCCGGGTTGCAGAGATGTGGGA
AGAGTTGCAGATTAAGGTGGATAAACTATCTAAGGCCTGATCTGAGGAGAGGTTTCGTTTA
ATGAGGAAGAAGAGCAGATTATCATCGACGTACATCGTATTCTTGGAACAAATGGGCTC
AGATTGCTAAGCACTTACCTGGACGCACTGATAATGAAGTCAAGAACTTTTGGAACATCAT
GCATTAGAAGAAACTCTTTCTCAAGGCTTAGATCCTTCTACACATAATCTTATGCCTT
CACACAAAAGATCTTCTTCTCAAACAATAATAATATCCCAAGCCAAACAAAACGACGT
CCATCATGAAGAACCCTACTGATCTTGATCAATCAACCACTGCTTTTCAATCACAACA
TCAATCCACCCACTTCCACTAAACCAAACTAAATCTCCTAACCAGACTACAATCC
CATCTCAAACCGTGATCCCTATCAATGATAACATGTCAAGTACTCAAACCATGATCCCTA
TCAATGATCCCATGTCAAGTCTTTTAGATGATGAGAATATGATTCCTCACTGGTCAGATG
TTGATGGAATGGCGATCCACGAAGCTCCGATGTTGCCTAGTGATAAGGCAGTAGTGGGAG
TGGATGATGATGATCTCAACATGGACATTTTGTGTTAACACTCCTTCTTCTGCTTTTG
ATCCTGATTTTGCTTCCATTTTCTCCTCTGCAATGTCTATCGATTTCAATCCCATGGATG
ATCTTGGCAGCTGGACCTTTTAGCTTTTACTCTACAGC

>G258 Amino Acid Sequence (domain in AA coordinates: 24-124)

MREKWEKMRDEMGRCCGKHVKRGLWSPEEDEKLLRYITTHGHPSSVSKLAGLQRCG
KSCRLRWINYLRLPDLRGGSFNEEEEQIIIDVHRILGNKWAQIAKHLPGRTDNEVKNFWNS

ACACTATGTAACGCTTGTGGAGTCCGGTTTAAATCCGGTAGACTTGTACCGGAATATAGA
CCGGCTTCGAGTCTACTTTTGTGTTTACTCAGCATTCAAACCTCTCACCAGAAAGTGATG
GAGCTTCGACGGCAGAAAGAAGTTATGAGACAACCACAACAAGTTCAACTTCATCACCAC
CACCACCGTTTTAG

>G343 Amino Acid Sequence (domain in AA coordinates: 178-214)
MDVYGLSSPDLLRIDDLDFSNEIDFSASSSGGSTAATSSSSFPFPQNPFSFHHHLPSSA
DHHSFLHDICVPSDDAAHLEWLSQFVDDSFADFPANPLGGTMTSVKTETSFPGKPRSKRS
RAPAPFAGTWSPMPLESEHQQLHSAAKFKPKKEQSGGGGGGGRHQSSSSETTEGGGMRR
CTHCASEKTPQWRTGPLGPKTLCNACGVRFKSGRLVPEYRPASSPTFVLQHSNSHRKVM
ELRRQKEVMRQPQQVQLHHHHHPF*

>G363 (1..780)
ATGAGACCAATATTAGACCTCGAAATTGAAGCTTCATCGGGCAGTAGTAGCAGCCAAGTG
GCCTCAAACCTTGCTCCGGTTGGGGAAGATTACAAACCAATCTCGCTGAATCTTAGCCTC
AGTTTCAACAACAACAACAATAATCTGGATCTTGAATCATCGTCTTTGACGCTGCCA
CTTTCGAGCAGAGTGAGAGTAGTAACCCGAGCAGCAGCAGCAACAACAACCATCTGTA
TCAAAGAGAGTCTTCTCTTGTAACTACTGCCAAAGGAAGTTCTATAGCTCTCAAGCGCTA
GGTGGTCACCAAAACGCTCACAACGTGAGAGAACACTCGCCAAACGCGCTATGCTATGG
GTCTTGCTGGGGTCTTCCCCGGTAGAGGATCAAGTAGCAATTATGCGGCTGCTGCCACAG
CAGCCGCTCTCGTGTGTCGCGCTTCACGGAAGCGGAAACGGGAACATGACATCGTTCAGG
ACTTTGGGAATCCGGGCACATTCTCGGCGCAGCAGCTCAGCATGACAAGGCAGACACCA
GAAACACTTATTAGAAAACATTGCCAGGTTCAACCAGGGGTATTTCCGTAATTGTATACCT
TTTTACGTGGAGGACGACGAGGCCGAGATGCTCTGGCCGGGAGTTTCCGGCAAGCTACG
AATGCGGTTGCGGTTGAAGCGGGTAATGATAATTTAGGTGAAAGAAAATGGATTCTTG
GACGTCAAGCAAGCGATGGATATGGAAGTTCTCTTCCAGATCTAACCTTGAAGCTTTGA
>G363 Amino Acid Sequence (domain in AA coordinates: 87-108)

MRPILDLEIEASSGSSSSQVASNLSVPVGEDYKPISLNLSLSFNNNNNNNLDLESSSLTLP
LSSTSESSNPEQQQQQPSVSKRVFSCNYCQRKFYSSQALGGHQNAHKRERTLAKRAMLW
VLLGSSPVEDQVAIMRLLPQQPLSCLPLHSGSGNGNMTSFRTLGIRAHSSAHDVSMTRQTP
ETLIRNIARFNQGYFGNCIPFYVEDDEAEMLWPGSFRQATNAVAVEAGNDNLGERKMDFL
DVKQAMDMESSLPDLTLKL*

>G370 (1..774)
ATGGACGAAACCAACGGACGAAGAGAACTCACGATTTTCATGAACGTCAACGTTGAATCC
TTCTCTCAGCTTCCTTTTATCCGCCGTACTCCTCCCAAAGAAAAGCCGCCATTATTCGT
CTCTTCGGCCAAGAGCTCGTCGGTGATAACTCCGACAACCTATCCGCAGAACCTTCTGAT
CATCAAACCACTACCAAGAACGATGAGAGCTCTGAGAATATCAAGGACAAAGACAAAGAA
AAAGATAAGGACAAAGACAAAGATAACAACAACAACAGGAGATTGAGTGTCCTACTGCT
TTCAGAAACTTCCCAACTTCTCAAGCCTAGGTGGACATCAAAACGCTCACAACGCTGAA
CGTCAACACGCCAAACGCGGTTCCATGACATCATACCTTCATCATCATCAGCCTCATGAC
CCTCACCACATCTACGGCTTCCTCAACAACCACCACCACCGTCACTATCCGTCTTGGACG
ACGGAAGCTAGATCATACTACGGCGGAGGGGACATCAAACGCCGTCGTACTACTCAAGG
AATACTCTTGCTCCTCCTTCTTAACCCACCGACAATCAACGGAAGTCCTTTAGGTTTG
TGGCGTGTAACCGCTTCCACGTCAACAAATACTATTCAAGGCGTTTACTCATCTTCACCA
GCTTCAGCGTTTAGGTGCGATGAGCAAGAGACTAATAAGGAGCCTAATAACTGGCCGTAC
AGATTGATGAAACCCAATGTGCAAGATCATGTGAGTCTCGATCTTCATCTCTGA

>G370 Amino Acid Sequence (domain in aa coordinates: 97-117)
MDEINGRRETHDFMNVNVEFSQLPFIIRTPPKKAAIIRLFQELVGDNSDNLSEPSD
HQTTTTKNDESSENIKDKDKDKDKDNNNNRRFECHYCFRNFPTSQALGGHQNAHKRE
RQHAKRGSMTSYLHHHQPHDHHIYGFLLNNHHHRHYPSTTEARSYYGGGGHQTPSYYSR
NTLAPPSSNPPTINGSPLGLWRVPPSTSTNTIQGVYSSSPASAFRSHEQETNKEPNWPY
RLMKPNVQDHVSLDLHL*

>G385 (37..2202)
TAGGGTTTGCTTTTCAGCTTTCCGGAGTATAAGAAAAGATGTTTCGAGCCAAATATGCTGCTT
GCGGCTATGAACAACGACGACAGCAATAACCACAACCTACAACCACGAAGACAACAATAAT
GAAGGATTTCTTCGGGACGATGAATTCGACAGTCCGAATACTAAATCGGGAAGTGAGAAT
CAAGAAGGAGGATCAGGAAACGACCAAGATCCTCTTCATCCTAACAAGAAGAAACGATAT
CATCGACACACCCAACCTTCAGATCCAGGAGATGGAAGCGTTCTTCAAAGAGTGTCTCTCAC
CCAGATGACAAGCAAAGGAAACAGCTAAGCCGTGAATTGAATTTGGAACCTCTTCAGGTC

AAATTCTGGTTCCAAAACAAACGTACCCAAATGAAGAATCATCACGAGCGGCATGAGAAC
TCACATCTTCGGGCGGAGAACGAAAAGCTTCGAAACGACAACCTAAGATATCGAGAGGCT
CTTGCAAATGCTTCGTGTCTTAATTGTGGTGGTCCAACAGCTATCGGAGAAATGTCATT
GACGAACACCAACTCCGTCTCGAAAATGCTCGATTAAAGGAAGAGATCGACCGTATATCC
GCAATCGCAGCTAAATACGTAGGCAAGCCAGTCTCAAACCTATCCACTTATGTCTCCTCCT
CCTCTTCCTCCACGTCCACTAGAACTCGCCATGGGAAATATTGGAGGAGAAGCTTATGGA
AACAAATCCAAACGATCTCCTTAAGTCCATCACTGCACCAACAGAATCTGACAAACCTGTC
ATCATCGACTTATCCGTGGCTGCAATGGAAGAGCTCATGAGGATGGTTCAAGTAGACGAG
CCTCTGTGGAAGAGTTTGGCTTTAGACGAAGAAGAATATGCAAGGACCTTTCCTAGAGGG
ATCGGACCTAGACCGGCTGGATATAGATCAGAAGCTTCGCGAGAAAGCGCGGTTGTGATC
ATGAATCATGTTAACTCGTTGAGATTCTCATGGATGTGAATCAATGGTCGACGATTTTC
GCGGGGATGGTTCTATAGCAATGACATTAGCGGTTTTATCGACAGGAGTTGCAGGAAAC
TATAATGGAGCTCTTCAAGTGATGAGCGCAGAGTTTCAAGTTCATCTCCATTAGTCCCA
ACACGTGAAACCTATTTTCGCACGTTACTGTAAACAACAAGGAGATGGTTCGTGGGCGGTT
GTCGATATTTCTGTTGGATAGTCTCCAACCAAATCCCCGGCTAGATGCAGGCGCGAGCT
TCAGGATGTTTGATTCAAGAATTGCCAAATGGATATTCTAAGGTGACTTGGGTGAGCAT
GTGGAAGTTGATGACAGAGGAGTTCACTAATACTATACAAACACATGGTTAGTACTGGTCAT
GCCTTCGGTGCTAAACGCTGGGTAGCCATTCTTGACCGCAATGCGAGCGGTTAGCTAGT
GTCATGGCTACAAAACATTTCTCTGAGAAAGTTGGCGTGATAACCAACCAAGAAGGGAGG
AGGAGTATGCTGAAATTTGGCAGAGCGGATGGTTATAAGCTTTTGTGTCAGGAGTGAGTGCT
TCAACCGCTCACACGTGGACTACATTGTCCGGTACAGGAGCTGAAGATGTTAGAGTGATG
ACTAGGAAGAGTGTTGGATGATCCAGGAAGGTCTCCTGGTATTGTTCTTAGTGACCCACT
TCTTTTTGGATCCCTGTTCTCTCAAAGCGAGTCTTTGACTTCTCAGAGACGAGAATTCA
AGAAATGAGTGGGATATTCTGTCTAATGGAGGAGTTGTGCAAGAAATGGCACATATTGCT
AACGGGAGGGATACCGGAAACTGTGTTCTCTTCTTCGGGTAAATAGTGCAAACCTCTAGC
CAGAGCAATATGCTGATCCTACAAGAGAGCTGCATTGATCCTACAGCTTCTTTGTGATC
TATGCTCCAGTCGATATTGTAGCTATGAACATAGTGCTTAATGGAGGTGATCCAGACTAT
GTGGCTCTGCTTCCATCAGGTTTTGCTATTCTTCTTGATGGTAATGCCAATAGTGAGGCC
CTGGGAGGAGATGAGGGGTGCTCTTGTACTGTTGCTTTTCAGATTCTGGTTGACTCAGTT
CCTACGGCTAAGCTGTCTCTTGGCTCTGTTGCAACTGTCAATAATCTAATAGCTTGCACT
GTTGAGAGAATCAAAGCTTCAATGTCTTGTGAGACTGCTTGAAAACCATCCATTAGC

>G385 Amino Acid Sequence (domain in AA coordinates: 60-123)
MFEPNMLLAAMNADSNHNHYNHEDNNNEGFLRDDEFDSPNTKSGSENQEGGSGNDQDPL
HPNKKKRYHRHTQLQIQEMEAFFKECPHPDDKQRKQLSRELNLPLQVQKFWFQNKRTQMK
NHHERHENSHLRAENELRNDNLRYREALANASCPNCGGPTAIGEMSFDEHQLRLENARL
REEIDRISAIAAKYVGKPVSNYPLMSPPLPPRPLELAMGNIGGEAYGNPNNDLLKSITA
PTESDKPVIIDLSVAAMEELMRMVQVDEPLWKSLLALDEEYARTFPRGIGPRPAGYRSEA
SRESAVVIMNHVNIVEILMDVNQWSTIFAGMVSRAMTLAVLSTGVAGNYNGALQVMSAEF
QVPSPLVPTRETYFARYCKQQGDGWSAVVDISLDSLQPNPPARCRRRASGCLIQELPNGY
SKVTWVEHVEVDDRGVHNLKHMVSTGHAFGAKRWVAILDRQCERLASVMATNISSGEVG
VITNQEGRRSMLKLAERMVISFCAGVSASTAHTWTTLSGTGAEDVRVMTRKSVDDPGRSP
GIVLSAATSFWIPVPPKRVDFDLRDENSREWDILSNGGVVQEMAHIANGRDITGNCVSL
RVNSANSQSNNMLILQESCIDPTASFVIYAPVDIVAMNIVLNGGDPDYVALLPSGFAILP
DGNANSGAPGGDGSLLTVAFQILVDSVPTAKLSLGSVATVNLIACCTVERIKASMSCET
A*

>G439 (128..967)
TATAAATCTTCGTTTCTACTTTTTTTTCTTCCATAATATAGTCAATTCTGTTTTCTTAATT
AGGGCTTCTTCTCTTGTCTTCTCCAATCTTATTAGTTTATTATTTATTTTGGTTATTG
TATACAAATGGCAATGGCTTTAAACATGAATGCTTACGTAGACGAGTTTCATGGAAGCTCT
TGAACCATTCTGAAGGTAACCTTCTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT
ATTAACCTCTAATTTCTATCCCTAATAATGACCAAGTCTTACCGGTATCTAACCACACCGG
TCCGATTGGGCTAAACAGCTCACTCCAACACAAATCCTCCAAATTCAGACAGAGTTACA
TCTCCGGCAAACCAATCTCGTCTGCTCGCTGGTAGTCATCTTCTCACCCTAAACCAAC
CTCAATGAAGAAAAATCGACGTAGCAACTAAACCGGTTAACTATACCGAGGCGTAAGACA
GAGGCAATGGGGTAAATGGGTAGCTGAGATTCTGGCTACCTAAAAACCGAACCCGGTTATG
GCTCGGTACGTTTCAACCGGCTCAAGAAGCTGCATTAGCTTACGATCAAGCAGCTCATAA
GATCAGAGGAGACAACGCTCGTCTCAATTTCCAGACATTGTTTCGTCAAGGACACTATAA

ACAGATATTGTCTCCGTCTATCAACGCAAAGATCGAATCCATCTGCAATAGTTCTGATCT
TCCACTGCCTCAGATCGAGAAACAGAAACAAAACAGAGGAGGTGCTCTCTGGTTTTTCCAA
ACCGGAGAAAGAACCGGAATTTGGGGAGATATACGGATGCGGATACTCGGGCTCATCTCC
TGAGTCGGATATAACGTTGTTGGATTTCTCAAGCGACTGTGTGAAAGAAGATGAGAGTTT
CTTGATGGGTTTGCAACAAGTATCCTTCTTTGGAGATTGATTGGGACGCTATAGAGAACT
CTTCTGAATCCATTTTATCTTTTTGATTCAATTTGTCTCTAAATGTAGAATTTTATTTTC
AGAGCTTTGTAAGGGAAGTTCTTGAATGAGAGTTGCAGAGGACTAGTGGAACCTAACTCT
GTTTTCTTTTGTAAGTATTGTTTATAATGGGCCGTTGAATGGCCTTATTGATTTAAACA
GCCCAAGTTTTTAAAAAAAAAAAAAAAAAAAAAAAAAAAA

>G439 Amino Acid Sequence (domain in AA coordinates: 110-177)
MAMALNMNAYVDFMEALEPFMKVTSSTSSNSTSNPKPLTPNFIPNNDQVLPVSNQTGPI
GLNQLTPTQILQIQTEHLRQNRQRRRAGSHLLTAKPTSMKKIDVATKPKLYRGVRQRQ
WGKWWABIRLPKNRTRLWLGTFFETAQEAALAYDQAAHKIRGDNARLNFDPDIVRQGHYKQI
LSPSINAKIESICNSSDLPLPQIEKQNKTEEVLSGFSKPEKEPEFGEIYCGYSGSSPES
DITLLDFSSDCVKEDESFLMGLHKYPSLEIDWDAIEKLF*

>G440 (237..1301)

AAAAAATCACTGTTTTCATAACACGTTTTTCTCTCTCACCCACCAAAAAAATCTTTTGT
TCTTGTACCAAAAAATCTCGTGATAAATCTCTTCAAACCTTGTGTTTATTTCTTCTTGA
TTCTCTCGAAATCTCTCTCAACAAACCCAGAACTTTCCTTGATTGCAAGCTTTTCTTC
CTTTTATATTCTTCATTTTGTATGCGAATATAGAGAGAGTCCATAAAGAAACAGTAATGG
ACGAATATATTGATTTCGACCATTTGAAGTACACAGAGCACAAGACTTCAATGACTAAAT
ACACCAAAAAAGTCATCGGAAAACTTTCCGGTGGTAAGTCATTGAAAAAGGTTAGTATTT
GTTATACTGATCCTGACGCAACAGATTCTCAAGTGACGAAGACGAAGAAGATTTCTTGT
TTCTCTCGCCGGAGAGTCAAAAGATTCTGTTAACGAGATCACTGTTGAGCCTAGCTGTAACA
ACGTCGTACCCGGAGTTTCGATGAAAGATAGAAAGAGACTCTCTTCTCCTCCGATGAAA
CTCAATCTCCGGCGTCGAGTCGTCAACGTCCTAATAACAAAGTTTCAGTCTCCGGTCAGA
TAAAGAAGTTCCGTGGTGTAGACAACGGCCATGGGGGAAATGGGCGGCGGAGATTAGAG
ATCCGGAGCAACGTCGGAGGATTGGCTCGGGACTTTTGAGACGGCGGAGGAAGCTGCCG
TGGTTTATGATAACGCCGCTATAAGACTCCGTGGACCGGACGCTTTAACTAATTTCTCCA
TACCGCCTCAAGAAGAGGAAGAAGAAGAACCAGGAAACCGGTTATTGAGGAGAAACCGG
TTATTATGACGACGCCAACACCAACAACATCGAGTTCTGAATCAACTGAAGAAGATTTAC
AACATCTCTCATCTCTACTTCCGTTCTCAATCACCGGTCAGAAGAGATTCAACAAGTAC
AACAACCGTTTAAATCAGCTAAACCCGAACCGGGGGTTTCAAATGCACCATGGTGGCATA
CCGGGTTTAAATACCGGTTTAGGTGAATCAGACGATTCAATTCCTTTGGATACTCCGTTTC
TTGACAACATTTCAATGAATCACCACCAGAGATGTCAATATTTGACCAACCAATGGATC
AAATTTCTGTGAAAATGATGATATCTTCAATGATATGTTGTTCTTGGGTGGTGAAACTA
TGAACATTGAAGATGAGTTAACAAGTTCTAGTATCAAAGATATGGGTTCAACGTTTAGTG
ATTTTGATGATTCAATTGATATCAGATCTATTAGTTGCTTAATATGATGATGAGAGTGAAG
AAGAAACCATCAAGCAAATATCTATGGTGTGACTGAAAAATTTGGTGTACTTTTTTTT
CTTTCATAAGTTCTATGAGCTTTTTTGTCTTTTTTTTTTAATAATTTATTTAGTTTGTCA
GGAGCTTGTAACAGTTTTTGGAGAAATAGTGGAATAATAGTTTAATTAATAAAAAAAAAA
AAAAAA

>G440 Amino Acid Sequence (domain in AA coordinates: 122-189)
MDEYIDFRPLKYTEHKTSMTKYTKKSSEKLSGGKSLKKVSICTDPDATDSSSEDEEDF
LFPRRRVKRFVNEITVEPSCNNVVTGVSMKDRKRLSSSDETQSPASSRQRPNNKVSVSG
QIKKFRGVRQRPWGKWAEEIRDPEQRRRIWLGTFFETAEEAAVVYDNAAIRLRGPDALTNF
SIPPQEEEEEEPEPVEIEKPVIMTTPPTSSSESTEEDLQHLSSPTSVLNHRSEEIQQ
VQPPFKSAKPEPGVSNAPWWHTGFNTGLGESDSSFPLDTPFLDNYFNESPPEMSIFDQPM
DQIFCENDDIFNDMLFLGGETMNIIEDELTSSSIKDMGSTFSDFDDSLISDLLVA*

>G5 (417..1421)

TTTTTTTTTTGCAATCTCCCCCTAATCTGTTGTTTCTCGCTTCTTCTTCTGTTAATCATC
TGTCTTTCAAAAAGAAAGAAAAAGAAAAATTCGATTTCTGGGTTTGTGTTTGTGCATACA
GAAAAAATCAAGCTTATGAATTTGTGTTTAAATTTTTGTTTAAATTTGAAAGGCAGGTT
TTTTCAGAACGAGATCGTTTTTTCAAATTTCTTCTGATTTTACCTCTTTTTTTCTTCTTA
GATTTTAGTGAATCGAGGGTGAATTTTTGATTCCTCTTTTCGGATCTACACAGAGGTT
GCTTATTTCAAACCTTTTAGATCCATTTTTTTTTTAATTTTCTCGGAAAAATCCCTGTTTC
TTTACTTTTTTATAAGTCTCAGGTTCAATTTTTTTCGGATTCAAATTTTTTAAATG

CAGCTGCTATGAATTTGTACACTTGTAGCAGATCGTTTCAAGACTCTGGTGGTGAAGTCA
TGGACGCGCTTGTACCTTTTATCAAAAGCGTTTCCGATTCTCCTTCTTCTTCTGCGAG
CGTCTGCGTCTGCGTTTCTTACCCCTCTGCGTTTTTCTCTCCCTCCTCTCCCGGTTATT
ACCCGGATTCAACGTTCTTGACCCAACCGTTTTTCATACGGGTCGGATCTTCAACAAACCG
GGTCATTAATCGGACTCAACAACCTCTCTTCTCTCAGATCCACCAGATCCAGTCTCAGA
TCCATCATCTCTTCTCCGACGCATCACAACAACAACCTTTTCTCGAATCTTCTCA
GCCCAAAGCGTTTACTGATGAAGCAATCTGGAGTCGCTGGATCTTGTTTCGCTTACGGTT
CAGGTGTTCTTTCGAAGCCGACGAAGCTTTACAGAGGTGTGAGGCAACGTCACCTGCGGAA
AATGGGTGGCTGAGATCCGTTTGCCGAGAAATCGGACTCGTCTCTGGCTTGGGACTTTTG
ACACGGCGGAGGAAGCTGCGTTGGCCATGATAAGGCGGCGTACAAGCTGCGCGCGGATT
TCGCCCCGGCTTAACCTTCCCTAACCTACGTCTAAGCGATTTCACATCGGAGGCGATTTCG
GTGAATATAAACCTCTTCACTCCTCAGTCGACGCTAAGCTTGAAGCTATTTGTAAAAGCA
TGGCGGAGACTCAGAAACAGGACAAATCGACGAAATCATCGAAGAAACGTGAGAAGAAGG
TTTCGTCGCCAGATCTATCGGAGAAAGTGAAGGCGGAGGAGAATTCGGTTTCGATCGGTG
GATCTCCACCGGTGACGGAGTTTGAAGAGTCCACCGCTGGATCTTCGCCGTTGTGCGACT
TGACGTTTCGCTGACCCGGAGGAGCCGCCGAGTGAACGAGACGTTCTCGTTGGAGAAGT
ATCCGTCGTACGAGATCGATTGGGATTCGATTCTAGCTTAGGGGCAAAATAGGAAATTC
GCCGCTTGCAATGGAGTTTGTGAAATTCGATGACTGGCCCAAGAGTAATTAATTAAT
ATGGATTAGTGTAAATTTTCGTATGTTAAATTTGTATTATGGTTTGTATTAGTCTCTCT
GTGTGCGGTCCAGCTTGGCGTTTTTTGTGTCAGGCTCGACCATGCCACAGTTTTTCATTTATG
TAATCTTTTTTTCTTTTGTCTTATGTAATTTGTAGCTTCAGTTTCTTCATCTATAATGCA
ATTTTATTATGATTATGTG

>G5 Amino Acid Sequence (domain in AA coordinates: 149-216)
MAAAMNLYTCSRSDSGGELMDALVPFIKSVSDSPSSSSAASAFSLHPSAFSLPPLPG
YYPDSTFLTQPFYSYSDLQQTGSLIGLNNLSSSQIHQIQSQIHHPLPPTHNNNNNSFSNL
LSPKPLLMKQSGVAGSCFAYGSGVPSKPTKLYRGVVRQRHWGKWVAEIRLPRNRTRLWLGT
FDTAEEAALAYDKAAYKLRGDFARLNFNLRHNGFHHGGDFGEYKPLHSSVDKLEAICK
SMAETQKQDKSTKSSKKREKKVSSPDLEKVKAEENSVSIGSPVPVTEFEESTAGSSPLS
DLTFADPEEPPQWNETFSLEKYPSYEIDWDSILA*

>G550 (1..1374)
ATGGCTGATCCGGCGATTAAAGCTCTTTGGAAGACGATTCCCTTTACCTGAGCTTGGTGTT
GTTGATCTTCTTCTAGCTATACCGGATTTTAAACCGAAACTCAGATTCTGTTTCGGTTA
TCAGATTTCGTGTACCGCGATGATGATGATGAAGAGATGGGTGATTCCGGTTTAGGACGA
GAAGAAGGTGATGATGTTGGTGATGGTGGAGGAGAGAGCGAGACTGATAAAAAGGAAGAA
AAGATAGTGAGTGTGAGGAAGAGTCATTGAGGAATGAATCTAATGATGTTACTACTACT
ACATCCGGGTATAACTGAAAAAACGGAACACAAAAGCTGCAAGACGAATGAAGAGTCA
GGTGGTACTGCTTGCTCTCAAGAGGGGAAGTTAAAGAAACCTGATAAGATTCTACCGTGT
CCGCGATGTAACAGCATGGAACCAAGTTCTGTACTACAACAATAATGTTAACC
CCTCGCATTTCTGCAAGAAATGTCAGAGATATTGGACAGCTGGTGAACGATGAGGAAT
GTTCCGGTTGGTGCTGGGAGACGTAAGAATAAGAGTCCAGCTTCTCATTATAACCGTCAT
GTAAGTATAACATCTGCCGAAGCTATGCAGAAGGTGGCGAGAAGTCTTCAACATCCT
AATGGTGCAATCTTCTCACTTTTGGCTCTGATCTGTGCTTTGTGAATCTATGGCTTCT
GGATTGAATCTTGTGAGAAGTCATTGTTGAAGACACAACTGTATTGCAAGAACCCAAT
GAAGGCTTGAAGATTACGGTTCCGTTAAACCAGACAAAAGAGCTGGAACAGTCCAGC
CCGTTACCAAAAAGTTCCATGCTTTCCAGGACCACCACCAACTTGGCCTTACGCTTGGAAC
GGAGTTTCGTGGACGATTTTACCGTTTTTACCCTCCACCGGCTTACTGGAGCTGCCCCGGG
GTTTACACCGGGGGCATGGAACAGCTTACATGGATGCCACAACCAATTACCATCTGGT
TCCAATCCAAATCTCTTACACTAGGTAAACATTCACGTGACGAGAACGCTGCTGAACCA
GGAACCGCTTTTGTATGAAACCGAGTCACCTTGGTAGGGAGAAAAGCAAACCGAGAGATGC
TTGTGGGTTCCCAAGACGCTGAGGATTGATGATCCAGAGGAAGCTGCTAAAAGTTCATC
TGGGAAACATTAGGGATCAAAAAGACGAAAATGCGGATACCTTTCGGAGCTTTTCAGATCA
TCAACCAAGAAAAAGCAGTCTTTCTGAAGGAAGACTTCCGGGAAGAAGACCGGAGTTG
CAAGCGAATCCTGCTGCTCTTTCTAGGTACGCAAACTTCCATGAGAGCTCATAG

>G550 Amino Acid Sequence (domain in AA coordinates: 134-180)
MADPAIKLFGKTIPLPELGVDSSSSYTGFLETQIPVRLSDSCTGDDDDDEEMGDSGLGR
EEGDDVGDGGGESETDKKEKDESECQEESLRNESNDVTTTTSGITEKTETTKAAKTNEES
GGTACSQEGKLLKPKDKILPCPRCNSMETKFCYNNYNNVNPQRHFCCKCQRYWTAGGTMRN

VPVAGARRKNKSPASHYNRHVSITSABAMQKVARTDLQHPNGANLLTFGSDSVLCESMAS
GLNLVEKSLKLTQTVLQEPNEGLKITVPLNQTNEEAGTVSPLPKVPCFPFPPTWPYAWN
GVSWTILPFYPPPAYWSCPGVSPGAWNSFTWMPQPNPSGSGNPNPSPTLGKHSRDENAAEP
GTAFDETESLGREKSKPERCLWVPKTLRIDDPBBAKSSIWETLGIKKDENADTFGAFRS
STKEKSSLSEGRLPGRRPQLQANPAALSRSANFHESS*

>G670 (28..1152)

CACAGCATTGCAGCTGTGAATAACTAAATGGGGAGACATTCTTGCTGTTACAAACAAAAG
CTGAGGAAAGGGCTTTGGTCTCCTGAAGAAGACGAGAAGCTTCTTACTCACATCACCAT
CACGGCCATGGCTGCTGGAGCTCTGTCCCTAAACTCGCTGGTTTGCAGAGATGTGGGAAG
AGTTGTGCGACTCGAGCAGATCTGGTACCGCCGACTAAGATGGATCAATTACTTGAGACCT
GATTTAAAGAGAGGAGCTTTTCTCCTGAAGAAGAGAATCTCATCGTCGAACCTTCATGCC
GTCCTTGGAACAGATGGTCACAGATTGCGTCAAGGCTTCCGGGTAGAACCACGACAACGAG
ATCAAGAATCTATGGAACCTCAAGCATCAAGAAGAACTGAAACAAAGAGGCATTGACCCA
AACACACACAAGCCCATCTCTGAAGTGGAGAGTTTATAGCGACAAAGACAAACCAACAACA
AGCAACAACAAAAGAAGCGGTAAACGATCACAAGTCTCCTAGTTCTCTTCTGCGACTAAC
CAAGACTTCTCCTCGAAAGGCCATCTGATTTATCCGACTACTTCGGATTTTCAAGAGCTT
AACTTCAACTCCAATCTAGGACTCTCTGTTACAACCTGATTCTTCACTCTGCTCGATGATT
CCGCCGAGTTTAGCCCCGGGAACATGGTTGGTTCTGTCTTCAGACACCAGTATGCGTA
AAGCCCTCGATTAGTCTTCTCCCGACAACAACAGTTCGAGTCTTATCTCCGAGGAGAT
CATGTGAAATTTGGCTGCACCAAACTGGGAATTTTCAAGACAAACAATAATACCTCAAAT
TTCTTCGACAATGGCGGATTTCTATGGTCTATCCCAAATTTCTTCTACTTCTTCTTCAAA
GTCAAACCAAATCATAACTTCGAAGAAATAAAATGGTCAGAGTATTTGAACACACCGTTC
TTCATAGGGAGTACTGTACAGAGTCAAACCTCTCAACCAATCTACATCAAATCAGAAACA
GATTACTTAGCCAATGTTTCAAACATGACAGATCCTTGGAGCCAAAACGAGAACTTGGGC
ACAACGTAACTAGTGACGTGTTCTCCAAGGATCTTCAGAGAATGGCCGTCTCTTTTGGT
CAGTCCCTTTAGCTTTTTTCTTTCTTTCTTTCTTATTTCTAACAGATGTAGAGAACATAA
AGATATACAAATACATACAATGTCAATACGTACAGTGGATTAAAGTGTCTGTATATTTT
ATGGGCGAGCTGTCTTTATTTTTATGTTTAAAAA

>G670 Amino Acid Sequence (domain in AA coordinates: 14-122)

MGRHSCCYKQLRKGLWSPEEDEKLLTHITNHGHGCVSSVPKLAGLQRCGKSCRLEQIWI
RRLRWINYLRPDLKRGAFSPREENLIVELHAVLGNRWSQIASRLPGRTDNEIKNLWNSSI
KKKLKQRGIDPNTHKPISEVESFSDKDKPTTSNNKRSNDHKSPPSSSATNQDFLERPS
DLSDYFGFQKLNFNNSNLGLSVTTDSSLCSMIPPQFSPGNMVGSVLQTPVCVKPISISLPPD
NNSSSPISGDHVKLAAPNWEFQTNMNTSNFFDNGGFSWSIPNSSTSSSQVKPNHNFEE
IKWSEYLNTPFFIGSTVQSQTSQPIYIKSETDYLANVSNMTPWSONENLGTETSDVFS
KDLQRMVAVSFGQSL*

>G760 (175..1878)

TGCTTAATTTCCAATGCCATCGTGATCGATTATCTCTCTCTCTCTCTTCCAATTTTCCCA
ATTCTTTTTTAAACCCCTAATTTTTCAGATATCTGATTATCTCTTGTATTTCTTCTACTC
GATTTGCTCCCATAAAAACCCCTTACTTTCTTCAAGTTCTGGTTTTCACCGATTGATGGGT
CGTGGCTCAGTGACGTGCTTGGTCTCTGGGTTCCGTTTTCACCCGACGGATGAGGAACTT
GTTGCTACTACCTTAAGCGTAAGGTCTGCAACAAACCCCTTAAGTTTCGATGCTATTTCC
GTCACCGACATATACAAGTCTGAGCCTTGGGATCTACCAGATAAGTCAAGCTGAAAAGT
AGAGACTTGGAAATGGTACTTCTTTAGTATGCTGGATAAGAAGTACAGTAATGGTTCCAAG
ACGAATCGTGCTACGGAGAAAGGGTATTGGAAGACGACTGGGAAAGATCGGGAGATTGCT
AATGGTTCAAGAGTCGTTGGGATGAAGAAGACACTTGTATTATCAAGGGTCGAGCTCCT
CGTGGTGAAAGGACCAATTGGGTTATGCATGAGTATCGGCTTTCTGATGAGGACTTGAAG
AAAGCTGGTGTGCCACAAGAAGCATATGTGTTATGTAGGATATTCAGAAAAGTGGTACG
GGTCTTAAGAAATGGGGAGCAGTATGTTGCTCCTTATCTTGAGGAGGAGTGGGAAGAAGAT
GGAATGACTTATGTACTGCTCAAGATGCTTTTCAAGTGAAGGATTGGCTTTGAATGATGAT
GTTTATGTCGATATTGATGACATTGACGAGAAGCCCGAAAATCTGGTGGTCTATGATGCC
GTTCTTATTTACCTAACTATTGTCATGGGGAATCAAGTAACAATGTTGAATCAGGCAAT
TACTCAGACTCTGGAAATTACATTCAACCAGGAACAATGTTGTCGACTCTGGTGGGTAC
TTTGAACAACCAATTGAAACTTTTGAAGGAAGATCGGAAGCCTATTATACGGGAGGGTAGC
ATTGAGCCTTGTCTCTGTTTCCAGAGGAACAATTTGGCTGTGGTGTGCAAGACGAAAAT
GTGGTGAATCTGGAATCTTCCAACAATAATGTGTTTGTAGCTGATACATGCTACAGTGAC
ATTCCTATTGATCATAACTATTTACCCGATGAGCCATTATGAGTATCCTAATAACAATCTT

CCACTCAACGATGGTCTGTACCTGGAACGAATGATCTCAGCTGTGCTCAACAAGATGAT
TTTAACCTCGAAGATTATCTCAGCTTCTTTGATGATGAGGGTTTGACTTTTGACGATTCT
CTATTAATGGGACCTGAAGATTTTCTTCCCAACCAAGAAGCCCTTGACCAGAAACCTGCC
CCTAAAGAATTGGAGAAGGAGGTGCGAGGAGGCAAGAGGCAGTGGAGGAAAAGGAAAGT
GGCGAAGGATCTTCTTCAAAACAAGATACAGATTTCAAGGACTTTGATTGAGCTCCGAAG
TACCCATTTCTCAAAAAGACGAGCCACATGCTTGGAGCCATTCTACTCCATCTTCATTT
GCTTCACAGTTCCAAACAAAGGACGCAATGCGTCTACACGAGCACAATCTTCTGGTTCA
GTTACAGTACTGACGGTATGATGAGAATATCAAACATGACTCTAGCAGCGGACAGCGGT
ATGGGCTGGTCATATGACAAGAACGGTAACCTCAACGTAGTCCTTTTCATTGGGGTAGTC
CAACAGGATGATGCGATGACTGCCTCGGGAAGCAAGACAGGAATTACGGCGACAAGAGCT
ATGTTAGTCTTCATGTGTTTATGGGTTCTCTACTCTCTGTTAGCTTCAAAATAGTAACC
ATGGTGTCTGCTCGGTAATAGGATCAAAGTTGAATCGTCTCAAAGACTTTTTTTGGTGT
TGTACCTCTCAATCATATAGCCTTTAACTTTGGCAGTGCTTTGCTGCTCAATATTTAAA
TTTTAAAAAAAAAAAAAAAAAA

>G760 Amino Acid Sequence (domain in AA coordinates: 12-156)
MGRGSVTS LAPGFRFHPTDEELVRYYLKRKVCNKPFKFD AISVTDIYKSEPWDL PDKSKL
KSRDLEWYFFSMLDKKYSNGSKTNRATEKGYWKTGKDREIRNGSRVGMKKTLLVYHKGR
APRGERTNWMHEYRLSDEDLKAGVPEAYVLCRIFQKSGTGPKNGEQYGAPYLEEEWE
EDGMTYVPAQDAFSEGLALNDDVYVDIDDIDEKPENLVYDVPILPNYCHGESSNNVES
GNYS DSGNYIQPGNNVVD SGGYFQPIETFEEDRKPIIREGSIQPCSLFPPEEQIGCGVQD
ENVVNLESSNNNVFVADTCYSDIPIDHNYLPDEPFMDPNNNLPLNDGLYLETNDLS CAQQ
DDFN FEDYLSFFDDEGLTFDDSLLMGPEDFLPNQEALDQKPAPKELEKEVAGGKEAVEEK
ESGEGSSSKQD TDFKDFDSAPKY PFLKKTSHMLGA IPTSSFASQFQTKDAMRLHAAQSS
GSVHV TAGMMRISNMTLAADSGMGWSYDKNGNLNVVLSFGVVQDDAMTASGSKTGITAT
RAMLVFMCLWVLLLSVSFKIVTMVSAR*

>G831 (92..1987)
TTCTTTTCATCGTGTGTCTATTATAAATATATGTCAATTGGTTTCTAAAAAATCTACC
ATTGATTGATTGATTTTTTTTTCTTTAAGAGATGAATTTATTTACAAGAATCTCATCTCG
GACTAAGAAGGCCAATCTTTACTACGTAACCCTAGTTGCTCTTCTCTGCATCGCTAGCTA
CCTTCTCGGTATTTGGCAAAACACGGCGGTAAATCCACGCGCCGCTTCGATGATTCAGA
CGGTACACCGTGCGAGGGATTACACAGACCTAATTCTACGAAAGATCTCGACTTCGACGC
GCATCACAACATTCAAGATCCACCTCCGGTGACGGAACCGCCGTTAGTTTCCCGTCGTG
TGCCCGCCGCTTGAGCGAGCACACGCCATGCGAAGACGCGAAGCGATCGTTGAAATCTC
GAGGGAGAGATTGGAGTATAGGCAAAGGCATTGTCCCGAGAGAGAAGAAATCTGAAGTG
CAGAATTCGGCGCCGTACGGTTACAAAACGCCGTTCCGATGGCCGCGGAGTCGTGACGT
GGCGTGGTTTCGCTAATGTGCCTCACACGGAGCTTACGGTTGAGAAAAAGAATCAGAATTG
GGTCCGGTACGAGAATGATCGGTTTGGTTCCCTGGTGGAGGTACGATGTTTCCACGTGG
CGCTGATGCTTACATTGATGATATCGGACGGTTGATTGATCTCAGCGACGGCTCTATCCG
TACAGCCATCGATACCGGTTGCGGGGTGGCTAGCTTCGGTGATATCTTTTATCAAGAAA
CATTACAACGATGTCAATTTGCACCAAGAGACACACAGAGCTCAAGTCCAGTTCGCACT
CGAGCGTGGTGTCCCGCGATGATCGGAATCATGGCTACAATCCGCCTACCGTACCCTTC
TAGAGCCTTTGATTAGCACATTGCTCTCGTTGCCTTATTCGGTGGGGCCAAAACGATGG
GGCTTACTTGATGGAGGTGGATAGGGTTTAAAGACAGGAGGGTACTGGATACTTTCTGG
ACCGCCGATTAATTGGCAGAAACGGTGGAAAGGGTGGGAACGGACCATGGATGATTGAA
TGCAGAGCAGACTCAGATCGAGCAGGTGCGGAGAAGCTTGTGTTGGAAGAAAGTTGTTCA
AAGAGATGATCTTGCTATTTGGCAAAAACCTTTAACCACATTGACTGTAAGAAAACAG
AGAGGTTTTGAAAAATCCGGAGTTTTGTGTCATGATCAAGATCCCGACATGGCCTGGTA
TACGAAGATGGATTETTTGTTTGACACCATTAACCTGAAGTTGATGACGCTGAGGATCTAAA
GACGGTGGCCGGAGGGAAGGTAGAAAAGTGGCCGGCTAGATTAAACGCGATTCTCCGAG
AGTAAACAAAGCGCTCTCGAGGAAATCACACCTGAAGCTTCTTGGAGAACACGAACT
GTGGAACAGAGAGTTTCTTATTACAAGAAGTTAGATTACCAAGTTGGGTGAAACCGGAG
ATACGAAACCTAGTCTGACATGAACGCTTACCTCGGTGGATTTCGCGCGGCTCTAGCGGA
TGATCCGGTCTGGTTCATGAACGTTGTCCCGGTGAGGCTAAGCTCAATACGCTCGGTGT
CATCTACGAGCGTGGTCTAATCGGAACGTATCAAACTGGTGTGAAGCCATGTGACGTA
TCCAAGAACGTATGATTTTATCCATGCTGACTCGGTTTTACATTGTACCAAGGTCAATG
TGAACCGGAGGAGATATTGTTGGAGATGACCGAATCTTAGACCGGGTGGTGGTGTGAT
TATAAGAGATGACGTGGACGTTTTGATCAAGGTTAAGGAATTAACCAAAGGATTAGAATG

GGAAGGTAGAATTGCTGACCACGAGAAGGGTCCTCATGAAAGAGAGAAGATTTACTATGC
GGTGAACAGTATTGGACCGTTCCTGCGCCTGATGAAGATAAAAACAACACTAGTGCTCT
CTCCTGATTTTTGAGTTTTTTTTTCTTACAATGTTTTTTTTTTTTTTCAATTTTT
TATACAACAATAAATTCTCAATAATTGTTGTGCGCGCCG

>G831 Amino Acid Sequence (domain in AA coordinates: 470-591)
MNLFTRISSRTKKANLYYVTLVALLCIA SYLLGIWQNTAVNPRAAFDDSDGTPCEGFTRP
NSTKDLDFDAHNNIQDPPPVETAVSFPSCAAALSEHTPCEDAKRSLKFSRERLEYRQRH
CPEREELKCRIPAPYGYKTPFRWPASRDVAWFANVPHTELTVEKKNQNWVRYENDRFWF
PGGGTMFPRGADAYIDDIGRLIDLSDGSIRTAIDTGCQVASFGAYLLSRNITMTSFAPRD
THEAQVQFALERGV PAMIGIMATIRLPYPSRAFDLAHCSRCLIPWGQNDGAYLMEVDRVL
RPGYIWILSGPFINWQKRWKGWERTMDDLNAEQTQIEQVARSLCWKKVVRDDLAIWQKP
FNHIDCKKTREVLKNPEFCRHDQDPDMAWYTKMDSCLTPLPEVDDAEDLKTAVAGGKVEKW
PARLNAIPPRVKNGALEEITPEAFLENTKLWKQSVSYKKLDYQLGETGRYRNLVDMNAY
LGGFAALADDPVWVMNVVPVEAKLNTLGVIYERGLIGTYQNWCEAMSTYPRTYDFIHAD
SVFTLYQGCQCEPBEILLEMDRILRPGGGVIIRDVDVLIKVKELTKGLEWEGRIADHEKG
PHEREKIYAVKQYWTVPAPDEDKNNTSALS*

>G864 (503..1534)
TGCAAAAACATTTTCTGTCTCTCCTCTGCCCAAATTTTTTTTCTTTCCAGGAATATTTT
CTAGAAAAACCAAGCAAAGCTTTAACCCCTTCTCCTCCAAAAGTAGCATCTTCTCTT
TTTTCTATTTCTCTTCTCTCTTCTATCTCTCTCTCGTTTGTGAACGATTCTTAAAGAT
ATAACCAAAAGCCCTTTTCTCTTTCTTCAACTTTCCGGGAAAAATCTTCACGCAGCAAG
GTTTCTCTCTCGGCTCTCGCAGTGTTTTTCCGGCCCTTTGTCTTTCTATAAAAAAAAAA
TTCGCGTCTCTTAAGAAAACTTTTTCCACCTAGAGAAGAAGAAGAGTATCACTCTTGTTG
TTCAAGTTTCTCTCTTTAATAAAAAATCCATCTTTATCTTTGTCTTCTTCTTTTGTG
TTTCCCTAATCTCTATGTTATAAACACACAGAGAGAAAAAGTCACAGTCTCGAGTCAA
AAACAGAGAATACGAAAGAAAAATGGAAGCGGAGAAGAAAAATGGTTCTACCGAGAATCAA
ATTACAGAGACAAAAACCAACGACAACAATCGTATCGGAGTTAACCAACACTCACC
AACCGGATTCTTCGTATCTCAGTCACTGACCCAGACGCTACTGATTCTCCAGTGACGA
CGAAGAAGAAGAACATCAACGCTTGTCTCTAAACGCCGTCGTGTTAAGAAGTTTGTCAA
CGAAGTCTATCTCGATTCCGGTGCTGTTGTTACTGGTAGTTGTGGTCAAATGGAGTCGAA
GAAGAGACAAAAGAGAGCGGTTAAATCGGAGTCTACTGTTTCTCCGTTGTTTCAGCGAC
GACGACTACGACGGGAGAGAAGAGTTCCGAGGAGTGAGACAGCGTCCATGGGGAAAATG
GGCGGCGGAGATAAGAGATCCGTTGAAACGTGTACGGCTCTGGTTAGGTACTTACAACAC
GGCGGAAGAAGCTGCTATGGTTTACGATAACGCCGCTATTAGCTTCGTGGTCCCGACGC
TCTGACTAATTTCTCAGTCACTCCGACAACAGCGACGGAAGAAGAAAGCCCAACCGTC
TCCGGTGAAGAAGAAGAAGAAAAACAACAAAAGCAAAAAATCCGTTACTGCTTCTTC
CTCCATCAGCAGAAGCAGCAGCAACGATTGTCTCTGCTCTCCGGTGTCTGTTCTCCGATC
TCCTTTCCGGTCGACGAATTTCCGGCATTCTTCATCACCAGTCGCGGCCGTTGTAGT
CAAGGAAGAGCCATCCATGACAACGGTATCTGAACTTTCTCTGATTTCTCGGCGCCCTT
GTTCTCAGATGATGACGTGTTTCGATTTCGGGAGCTCAGTGGTTCCCGACTATCTCGGCGG
CGATTTATTTGGGGAAGATCTATTACGGCGGATATGTGTACGGATATGAACCTTCGGATT
CGATTTCCGATCCGATTATCCAGCTGGCACATGGAGGACCATTTCAAGATATCGGGGA
TCTATTCCGGTCCGATCCTCTTTAGCTGTTTAAATAATTTTAAATAAATAAATAGTTA
TACCGGCCGTTACTAAACGGAACCGGAGAAAGTTTTGTATACCGGTGACATAAAATCTCG
GTTATGTTTCGTAATCTTTTTTCTTTGTTATATATAAAAAATGAATGAACTGAATTAA
TGTAAGTTAATGGTGATAATTATTAACGTTTTAAGTTTTGAAAAAAAAAAAAAAAAAAAA
AAAAAA

>G864 Amino Acid Sequence (domain in AA coordinates: 119-186)
MEAEKKMVLPRIKFTEHKINTTIVSELTNTHQTRILRISVTDPDATDSSSDDEEEHQR
FVSKRRRVKFKVNEVYLDGAVVTGSCGQMESKKRQKRAVKSESTVSPVVSATTTTGEK
KFRGVRQRPWGKWAAEIRDPLKRVRLWLGTYNATAEAAAMVYDAAIQLRGPDALTNFSVT
PPTATEKKAPPPSPVKKKKKKNNKSKSVTASSISRSSNDCLSPSVSLRSPFAVDEF
SGISSSPVAVVVKEEPSMTTVSETFSDFSAPLFSDDVDVDFRSSVVPDYLGGDLFGEDL
FTADMCTDMNFGDFGSLSSWHMEDHFQDIGDLFGSDPLLAV*

>G884 (31..1575)
TTTTTTTTTGTGTTAATTTTGGGGATCGATGTGGGAAAAGGAAGAAGCTCCGTCGACA
TCGAAGTCCACCGGAGCTCCGTCGCGTCCGACTTTATCTCTTCTCCACGGCCGTTTAGT

GAGATGTTCTTTAACGGTGGCGTTGGATTCAAGTCCTGGTCCGATGACTCTGGTCTCTAAT
ATGTTCCCTGATTCCGATGAGTTTAGGTCTTCTCTCAGCTTCTCGCTGGAGCCATGTCT
TCTCCAGCGACTGCAGCTGCTGCTGCTGCTGCTGCGACGGCTAGTGATTACCAGAGACTT
GGTGAAGGGACTAATAGTCTTAGTGGTGATGTTGACCCGAGATTCAAGCAAAACAGACCA
ACCGGTTTGGATGATTTCTCAATCTCAATCGCCGTCGATGTTACCCGTACCGCTGGTTTA
AGTCCAGCTATGTTGCTCGATTACCAAGCTTTTTGGGTCTTTTCTCTCCCGTTCAGGGA
TCATATGGAATGACACATCAGCAAGCTCTAGCTCAAGTCACTGCTCAAGCAGTTCAAGCC
AATGCCAATATGCAACCACAAACAGAGTACCTCCTCCCTCTCAAGTTCAATCATTTTCA
TCGGGTCAAGCGCAGATCCCGACCTCGGCTCCACTACCAGCTCAAAGAGAAACCTCAGAT
GTAACCATCATAGAGCACAGGTCAACAACAGCCTCTAAATGTTGACAAACAGCTGATGAT
GGCTATAACTGGCGAAAATATGGGCAAAAGCAAGTTAAAGGTAGCGAGTTTCCACGAAGC
TATTACAAGTGTACTAATCCAGGATGTCTGTCAAGAAGAAGGTTGAGAGATCTCTTGAT
GGACAAGTAACGGAGATTATCTACAAAGGTCAGCACAATCATGAACCTCCTCAAAACACT
AAGCGAGGTAACAAAGATAACACCGCGAATATAAATGGGAGTTCGATAAATAACAATCGC
GGGAGTTCTGAATTGGGGGCATCAGTTTTCAAATAATAGCTCCAACAAGACTAAGAGA
GAGCAACATGAAGCAGTAAGTCAAGCTACGACAACAGAGCACTTGTCTGAGGCAAGTGAC
GGTGAAGAAGTTGGTAATGGAGAAACTGATGTGAGAGAGAAAGATGAGAATGAGCCTGAT
CCCAAGAGAAGAAGTACAGAAGTTCCGATTTCAGAACCAGCTCCTGCTGCTTCACATAGA
ACTGTGACAGAGCCTAGAATTATTGTCCAAACGACGAGTGAAGTTGATCTTCTAGATGAT
GGATATAGGTGGCGTAAATATGGACAGAAAGTTGTCAAAGGAATCCTTATCCGAGGAGC
TACTACAAGTGCACAACACCAGGATGTGGTGTGAGGAAACATGTAGAGAGAGCAGCAACA
GATCCAAAAGCTGTAGTTAACAACATATGAAGGAAAACATAACCATGACCTTCCCGCTGCT
AAATCAAGCAGCCATGCCGCTGCAGCGGCACAGTTAAGGCCAGATAATCGACCTGGCGGT
TTGGCTAACTTAAATCAACAGCAGCAGCAACAGCCCGTTGCGCGGCTAAGGCTTAAAGAA
GAGCAAACTTGAGAGAAGAAAACCTTTGACCGTTTTTCATTACAAAAGCTTTCAAAT
TCCACTCACACACTTGTCTGAAAAATCTAGCAGTTTGCAGGAAAGAAACAGCTTCAAGAG
GTTGTAGTTCTTCTATGTTCTGGTGTAAAACTTAAAGCTTTTAGGGTTTTCAGATTTT
TGTTTACTAATACTGTATGTGAATCTTTTGTACATGAGGAAGAAAATTACAGGGGGATA
TTTTGTGTTGTATCTTTTGTGTTATGTTTCAGTAAAGATAGGTCTTACATTTTGTGTA
AAAAAAAAAAAAAAAAAAAA

>G884 Amino Acid Sequence (conserved domain in AA coordinates:227-285, 407-465)

MSEKEEAPSTSKSTGAPSRPTLSLPPRPFSEMFNGGVGFSPGPMTLVSNMFPDSDEFRS
FSQLLAGAMSSPATAAAAAAAAAATASDYQRLGEGTNSSSGDVDPFRFKQNRPTGLMISQSQS
PSMFTVPPGLSPAMLLDSPSFLGLFSPVQGSYGMTHQQLAQVTAQAVQANANMQPQTEY
PPPSQVQSFSSGQAQIPTSAPLPAQRETSVDVTIIIEHRSQQPLNVDKPADDGYNWRKYGQK
QVKGSEFPFRSYKYCTNPGCPVKKKVERSLDGQVTEIIYKGQHNHEPPQNTKRGNKONTAN
INGSSINNRRGSSELGASQFQTNSSNKTREQHEAVSQATTTEHLSEASDGEEVGNGETD
VREKDENEPDPKRRSTEVRISEPAPAASHRTVTEPRIIVQTTSEVDLLDDGYRWRKYGQK
VVKGNPYPRSYKYCTTPGCGVRKHVERAATDPKAVVTTYEGKHNHDLPAKSSSHAAAAA
QLRPDNRPGGLANLNQQQQQQPVARLRLKEEQTT*

>G898 (161..772)

GAAGAAAAGATTCAAAAACCTAGATTTCACAAAATCGATTGGCTGTCAAATTTCTCTCC
GGCGATTTTCTCTGAGTGAAATTCGGCTCAAGGTGATTATAGCGATCATCGAATCAAATT
GATTGAAGAGGTACAAAGGTAGTTACTTTGAGCTGAAAGATGAACACGTGAGAGGTGAG
AGTACCTCGAGGAAATCGACGGAGGAAAGCTGTGATTGATCTGAATGCGGTACCTGTTGA
TCAAGAAGGGACCTCTGCTTCTGTTAGAACTCTTACGGTGCCTATTACACCGTCTCAGCC
TGCTCCTACGATGATTGATGTCGATGCTATTGAGGATGATGTTATTGAATCATCCGCTAG
TGCTTTTGTCTGAAGCTAAAAGCAAATCAAGAAATGCACGTCGGAGACCTTTGATGGTTGA
TGTAGAGTCAGGAGGTACGACTAGATTCCCTGCCAACATAAGCAACAAACGCAGAAGGAT
TCTTCTAGTGAATCTGTCTGACTGTGAGCATGCCTCTGTAAATGATGAAGTCAACAT
GTCTTCGAGAGTGTCTAGATCAAAGGCTCCAGCTCCTCCACCAGAAGAGCCAAAGTTTAC
ATGTCCAATCTGCATGTGTCCCTTTACGGAGGAGATGTCAACCAAGTGCAGGTACATCTT
CTGCAAGGATGTATAAAGATGGCAATATCTCGCCAGGGCAAATGCCCTACTTGTAGGAA
AAAGGTTACTGCAAAAGAGCTGATTTCAGTTTTCTTCCAACCACTAGATGAGTGGTCCG
GCAACATCACCAGCCACCCTGTCTAATGGTTTATCAGACTATCCTCCTATTCACTTTGGA
ACATTGAAGGGACTTCGTTGACTTGGTATTTTGAATATTTTGTCTTGTGGAAGAGAAA
TATTCAGTGATCAAGAAGCCAGAAGGCCCTATCATTGATGGATATCATTTGGTAATAACT

CTTTGTTTTTAGTTGTTGTTCTATGTAATTTAGGTCTCTGCAAACCTCTCAGTCGATACT
CTTCTCTCTTGATAGATGATAAGATATATGGAAAAAAATTAATATTGAATCTTTACTA
AAA

>G898 Amino Acid Sequence (domain in AA coordinates: 148-185)
MNTSEVRVPRGNRRRKAVIDLNAVVDQEGTSASVRTLTVPITPSQPAPTMIDVDAIEDD
VIESSASAFAEAKSKSRNARRRPLMVDVESGGTTRFPANISNKRRRIPSSSEVIDCEHAS
VNDEVNMSRRVSRSKAPAPPPEEPKFTCPICMCPFTEEMSTKCGHIFCKGCIKMAISRQG
KCPTCRKKVTAKELIRVFLPTTR*

>G900 (1..648)
ATGGGGGAAGAAGAAGTGCAGATTATGTTGTGGTGTAGCGAGAATGTATTGTGAGTCAGAT
CAAGCGAGTTTATGTTGGGATTGTGACGGTAAAGTTCACGGAGCTAATTTTCTGGTGGCG
AAACACATGCGTTGTCTTCTATGTAGCGCGTGTGAGTCACACACGCCCTTGAAAGCTTCT
GGGCTGAATCTTGCGCCAACTGTTTCTATCTGTGAGTCTGTTTAGCTCGTAAGAAGAAT
AACAACAGCTCCCTCGCCGGGAGGGATCAGAATCTTAACCAAGAAGAAGAGATCATTGGT
TGTAACGACGGAGCTGAGTCTTATGATGAGGAAAGCGATGAGGATGAAGAAGAAGAAGAA
GTGGAGAATCAGGTTGTTCCGGCTGCGGTGGAGCAAGAACTTCCGGTGGTGAAGTTCGTGCG
TCTTCCGTTAGTAGTGGTGAAGGAGATCAGGTGGTGAAAAGGACGAGACTTGATTTGGAT
CTTAACCTCTCCGATGAGGAGAACCAATCTAGACCATTTGAAAAGATTATCGAGAGACGAA
GGTTTGTCAAGATCAACTGTTGTGATGAATAGCTCAATCGTGAAATTACACGGAGGGAGG
AGAAAAGCAGAGGGATGTGATACATCATCGTCGTCTTCTGTTTATTGA

>G900 Amino Acid Sequence (domain in AA coordinates: 6-28, 48-74)
MGKKKCELCGVARMYCESDQASLCWDCDGKVHGANFLVAKHMRCLLSACQSHTPWKAS
GLNLGPTVSICESCLARKKNNSSLAGRDQNLNQEIIIICNDGAESYDESEDEDEEEEEE
VENQVPPAAVEQELPVVSSSSSVSSGEGDQVVKRTRLDLDLNLSDENQSRPLKRLSRDE
GLSRSTVVMNSSIVKLHGRRKAEGCDTSSSSSFY*

>G913 (108..806)
CATTCAAAAACATCATATATATACACAAACACACTTTGATACAACAAAAAACAGAAC
ACAAACAAAAACACATTGTAACATTAGTTTAAAGCATTAAAGCTTCTTTATGTGAATAATA
ATAATTCTCCGACCACCGTGAATCAAGAAACGACGACGTCTCGTGAAGTCTCAATCACAT
TGCCTACTGATCAATCTCCTCAAACCTCACCAGGATCATCTTCTCTCCTTACCAGAGAC
CTTCCGGTGGATCACCGGCGAGAAGAACGGCGACTGGATTATCCGGCAAGCACTCTATTT
TCAGGGGGGATTGCACTACGTAACGGAAAATGGGTATCGGAGATTAGAGAGCCACGTAAAA
CGACAAGAATTTGGCTCGGGACTTATCCGGTACCGGAGATGGCTGCCGCCGCTTACGACG
TGGCTGCGTTAGCTTTAAAGGACCCGACGCCGTTTTGAATTTTCTTGGTTTAGCTTTGA
CTTACGTGGCTCCGTTTCAAACCTCTGCTGCGGATATAAGAGCGGCTGCTAGTAGAGCAG
CGGAGATGAAGCAACCGGATCAGGGTGGGGATGAGAAGGTATTGGAACCGTTCAACCCG
GCAAAGAGGAAGAATTAGAAGAAGTGCTGTAACCTCGTTCGTTGGAGTTTATGGATG
AGGAAGCGATGTTGAATATGCCGACTTGTGTTGACGGAGATGGCTGAAGGGATGTTGATGA
GTCCACCGAGAATGATGATACATCCGACGATGGAAGATGATTCGCCGGAGAATCATGAAG
GAGATAATCTTTGGAGTTATAAATGAATCCATTGAAGCTGCTCTCTTTTTTATGTTTTTC
CGGTGCAATGAGATTTTCCCCCTTTTFTTTTTTCTTTTTGGGTGCGTGT

>G913 Amino Acid Sequence (domain in AA coordinates: 62-128)
MSNNNNSPTTVNQETTTTSREVSTLPTDQSPQTSPGSSSPSPRPSGGSPARRTATGLSG
KHSIFRGLRLNGKWVSEIREPRKTTRIWLGTYPVPEMAAAAYDVAALALKGPDVNLNFP
GLALTYVAPVNSAADIRAAASRAEMKQPDQGGDEKVLPEVPQPKKEELEEVCNSCSL
EFMDEEAMLNMPILLTEMAEGMLMSPPRMMIHPTMEDDSPENHEGDNLSYK*

>G937 (45..1046)
TGGAAAAAGTTTGAATTTTAAATTCGAATCGAGAAAAAATAAAAAATGGGTTCTTTAGGTG
ATGAGCTTAGTTTGGGATCGATCTTTGGGAGAGGAGTTTCGATGAATGTTGTGGCGGTTG
AGAAAGTTGATGAACATGTTAAGAAGCTTGAAGAAGAGAAGAGAAAGCTCGAAAGTTGTC
AAGTTGAGCTTCCTCTGCTTTTGCAGATTTTAAACGATGCGATTTTGTATCTGAAGGATA
AGAGATGTTTCAGAGATGGAGACTCAACCATTTGTTGAAAGATTTCAATTTCTGTTAATAAAC
CTATTCAAGGAGAAAGAGGAATAGAATTGCTGAAAAGAGAGGAGCTAATGAGGGAGAAGA
AGTTTCAGCAATGGAAGCTAATGATGATCACACTAGTAAGATCAAGAGCAAGCTTGAGA
TTAAGAGAAATGAGGAGAAATCTCCTATGTTGTTGATTCCAAAGGTGGAACTGGTTTAG
GCCTCGGTTTAAAGTTTCGAGTTTCGATAAGAAGAAAAGGGATTGTTGCCTCATGTGGCTTTA
CTTCTAATCTATGCCACAACCAACCAACACAGCAGTACCACAACAACAGCATTCTTCTTA

AGCAGCAAGCTTTACGGAAGCAAAGAAGGTGTTGGAATCCAGAGTTGCATCGCCGATTG
TCGATGCATTGCAACAGCTAGGTGGACCGGGAGTGGCAACTCCTAAACAAATTAGAGAAC
ATATGCAAGAAGAAGGCTTAACCAATGATGAAGTCAAGAGTCATTTACAGAAATACAGGT
TACACATCAGGAAGCCAAATTGCAATGCGGAGAAACAATCAGCAGTTGTTTTAGGGTTTA
ACTTGTGGAAATCTCTCAGCACAAAGATGAAGAAGAGACATGTGAAGGAGGAGAATCATTGA
AGAGAAGCAATGCGCAATCAGATTCTCCTCAAGGTCCTTTGCAGTTACCGTCTACAACAA
CAACAACTGGTGGAGATAGTAGCATGGAAGATGTTGAAGATGCTAAGTCTGAGAGCTTTC
AACTGGAGAGATTGAGATCACCATAAATCTCAAGAAACCAAACTCTTGATCACGGTTTTG
TTATTTTGGATTCATTACTATATCTATTAGTAGTGAATGAGAACATAATTATAGAAAGG
TTTATAGATATATATATAGAGAAAAAGAGAGAGTGAGGATGGTTCAAATTATTTGCAGA
>G937 Amino Acid Sequence (conserved domain in AA coordinates: 197-246)
MGS LGDELSLGSIFGRGVSMNVVAVEKVD EHVKKLEEEKRKLESCQLELPLSLQILNDAI
LYLKD KRCSEMETQPLLLKDFISV NKPIQGERGIELL KREELMREKKFQQWKANDDHTSKI
KSKLEIKRNEEKSPMLLIPK VETGLGLGLSSSSIRRKGI VASCGFTS NSMPQPPTPAVPQ
QPAFLKQQALRKQRR CWNPELHRRFVDALQQLGGPGVATPKQIREHMQEGLTNDEVKSH
LQKYRLHIRKPN SNAEKQSAVVLGFNLWNSSAQDEEETCEGGESLKRSNAQSDSPQG PLQ
LPSTTTTTTG DSSMEDVEDAKSES FQLERLRSP*
>G960 (63..1538)
TACCGTCGACCCACGCGTCCGAGTG TATTCAAAGTCGGAAGAAACCTAAAGAAGAGGA
TTATGGGTGCTGTATCGATGGAGTCGCTTCCTTTAGGTTTCAGATTACAGCCTACCGATG
AAGAGCTCGTCAATCACTACCTCCGCTCTCAAGATCAACGACGTCACCTCCGATGTCCGTG
TCATCCCTGATATCGATGTCTGCAAATGGGAACCTTGGGATCTTCTGCTCTCTCGGTGA
TTAAGACGGATGATCCAGAGTGGTTCTTTTTCTGCCCTCGTGATCGGAAATACCTAATG
GTCATCGCTCTAACAGAGCAACTGACTCTGGCTATTGGAAGCTACTGGTAAAGATCGTA
GCATCAAGTCTAAGAAGACTTTAATCGGTATGAAGAAGACTCTTGCTCTTATCGTGGAC
GAGCTCCTAAAGGTGAGCGGACTAATTGGATTATGCACGAGTATCGTCCCACTCTTAAGG
ATCTTGATGGCACTTCCCCTGGCCAAAGCCCTTACGTTCTTTGTGCGCTCTTCCACAAGC
CTGATGATCGGGTTAATGGTGTCAAGTCCGATGAAGCAGCTTTTACGGCCAGCAACAAAT
ACTCACCTGATGATACATCATCTGATCTTGTTCAGAAACACCTTCCTCTGATGCTGCTG
TTGAGAAACCATCAGATTATTCAGGTGGATGCGGTTATGCTCATAGTAATAGTACCGCAG
ATGGGACAATGATTGAGGCACCTGAAGAGAATCTTTGGTTATCTTGTGACCTTGAAGATC
AAAAGGCACCACTACCGTGTATGGATTCTATATATGCTGGTGATTTCAATTACGATGAGA
TTGGATTCCAATTTCAAGATGGTACCAGCGAACCAGATGTATCACTAACAGAATTGTTGG
AGGAGGTGTTCAATAACCCCTGATGACTTCTCTTGCGAGGAATCGATCAGTCGAGAGAATC
CAGCAGTCTCACCAAATGGGATATTTTCATCTGCTAAAATGCTGCAGTCTGCAGCACCAG
AGGATGCTTTCTTCAACGACTTCATGGCTTTCACTGATACAGATGCTGAGATGGCGCAAT
TGCAGTATGGTTCAAGAGGTGGAGCTTCTGGTTGGCCAAAGTGACACTAATTCATACTATA
GTGATTTGGTTCCAGCAAGAGCAAATGATCAATCATAACACAGAGAACAACCTCACAGAAG
GGAGAGGGATAAAGATCCGGGCTCGACAGCCTCAGAACCGGCAGAGTACAGGATTGATAA
ACCAGGGTATTGCTCCAAGGAGAATCCGTCTGCAGCTGCAGTCTAACTCTGAAGTAAAAG
AACGAGAGGAGGTGAATGAAGGACACACTGTTATTCCCGAGGCCAAAGAAGCTGCAGCTA
AATACTCAGAGAAGAGTGGTTCTTTGGTTAAACCTCAAATAAAGCTCAGGGCGCGGGGAA
CTATAGGCCAAGTAAAAGGAGAGAGATTTCAGACGACGAGGTACAGGTGCAGAGCACAA
AGAGAGAGAGAGAGAGAATCAAATGTAGTTTAATGTAATTAGGGATGATGCAATGTTAGC
ATGTTTGTGTGTTGTAACCTTAAAACTTATTTAGGAATCTGATAAAAGTTACTGTTGAAA
AAAGAAAAA AAAAAAAAAAAAAAAAAAAAAA
>G960 Amino Acid Sequence (domain in AA coordinates: 13-156)
MGAVSMESLPLGFRFRPTDEELVNHYLR LKINGRHSDVRVIPDIDVCKWEPWDLPALSVI
KTDDPEWFFFCPRDRKYPNGHRSNRATDSGYWKATGKDRS IKS KTLIGMKTLV FYRGR
APKGERTNWIMHEYRPTLKDLDGTS PGQSPYVLCRL FHKPDDRNVGVKSDEAAFTASNKY
SPDDTSSDLVQETPSSDAAVEKPSDYSGGCGYAHNSN STADGMTIEAPEENLWLSCDLEDQ
KAPLPCMDSIYAGDFS YDEIGFQFQDGTSEPDVSLTE LLEEVFNPNDDFSCEBSISREN P
AVSPNGEIFSSAKMLQSAAPEDAFFNDFMAFTD TDAEMAQLQYGEAGGASGWPSD TNSYYS
DLVQQEQMINHNTENNLT EGRGIKIRARQPQNRQSTGLINQGIAPRRIRLQLQSNSEVKE
REEVNEGHTV IPEAKEAAKYSEKSGSLVKPQIKLRARGTIGQVKGERFADDEVQVQSTK
RERERIKCSLM*
>G991 (6..533)

GAAAAATGGAAGAAGAAAAGAGATTGGAGCTAAGGCTAGCTCCTCCTTGTCACCAATTCA
CTTCCAACAACAATCAATGGATCTAAACAAAAAGCTCGACCAAGAAACATCATTC
TTTCCAATAACAGGGTTGAGGTAGCTCCAGTGGTGGGATGGCCCGGTGAGATCATCCC
GGAGAAACCTAACGGCACAATAAAGGAGGAGATGAAGAAGAAGGAGAGTGATGAAGAGA
AGGAATTGTACGTTAAGATCAACATGGAAGGAGTTCCAATAGGAAGAAAAGTCAACCTTT
CAGCTTTATAACAACCTACCAACAGCTTTTACATGCCGTTGACCAACTCTTCTCTAAGAAAAG
ATTCGTGGGATCTAAACAGACAATACACTTTGGTCTACGAAGACACTGAAGGAGATAAAG
TTCTGGTCGGGGATGTTCTTGGGAGATGTTTGTATCTACTGTAAAGAGGTTGCATGTTT
TAAAGACCTCCCACGCCTTCTCACTCTCACCTAGAAAACATGGCAAGGAATAGAGAGAGG
TTGGCCAAAATCATCAGTTCGATGGTTTGTTTTAAATGTAATTTTGTGGAAACTAATGG
GGTTTGGCTTTGATTTACTGGTTTTCTTTTCACTTATGTACTAGGTTTTTGTCTTGCTAT
GTTATTTCTTGTTTTGGTTGTAAATATGCTGTTTTCGTTTAAAGAAATCGGGGGTTAGTATGT
TATCGTGTGTATAAAAAATAGTGAAGCACGTAAAGTTGATTACAAAAA

AAAAA

>G991 Amino Acid Sequence (domain in AA coordinates: 7-14,48-59,82-115,128-164)

MEEEKRLELRAPPCHQFTSNNNINGSKQKSSTKETSFLSNRVEVAPVVGWPPVRSSRR
NLTAQLKEEMKKKESDEEKELYVKINMEGVPIGRKVNLSAYNNYQQLSHAVDQLFSKKDS
WDLNRQYTLVYEDTEGDKVLVGDVPWEMFVSTVKRLHVLKTSFAFSLSPRKHGKE*

>G748 (98..1444)

CCACGCGTCCGCACTCTCCCAAATCTCTCTTCTTAAACAACAAAAAATCACAGAGA
CATAGAGAGAAGAAGACGGAACAGAGGCTCCAAAAAATGATGATGGAGACTAGAGATCC
AGCTATTAAGCTTTTCGGTATGAAAATCCCTTTTCCGTCGGTTTTTGAATCGGCAGTTAC
GGTGGAGGATGCAAGAAGATGACTGGAGCGCGGAGATGACAAATCACCAGAGAAGGT
AACTCCAGAGTTATCAGATAAGAACAACAACACTGTAACGACAACAGTTTAAACAATTC
GAAACCCGAAACCTTGGACAAAGAGGAAGCGACATCAACTGATCAGATAGAGAGTAGTGA
CACGCTGAGGATAATCAGCAGACGACACCTGATGTTAAACCTTAAAGAAACCGACTAA
GATTCTACCGTGTCCGAGATGCAAAAGCATGGAGACCAAGTTCTGTTATTACAACAATA
CAACATAAAACCGCCTCGTCATTCTGCAAGGCTTGTGAGAGATATTGGACTGCTGGAGG
GACTATGAGGAATGTTCTGTGGGGGACGACGTCGTAAGAACAAGCTCATCTTCTCA
TTACCGTCACATCACTATTTCGAGGCTCTTGAGGCTGCGAGGCTTGACCCGGGCTTACA
GGCAAAACACAAGGGTCTTGAGTTTGGTCTCGAAGCTCAGCAGCAGCAGTTGCTGCTCC
CATGACACCTGTTATGAAGCTACAAGAAGATCAAAGGTCTCAAACGGTGCTAGGAACAG
GTTTCACGGGTTAGCGGATCAACGGCTTGTAGCTCGGGTAGAGAATGGAGATGATTGCTC
AAGCGGATCTCTGTGACCACCTTAACAATCACTCAGTGGATGAATCAAGAGCACAAAG
CGGCAGTGTGTTGAAGCACAAATGAACAACAACAACAATAACATGAATGGTTATGC
TTGCATCCCAGGTGTTCCATGGCCTTACACGTGGAATCCAGCGATGCCCTCCACCAGGTTT
TTACCCGCTCCAGGGTATCCAATGCCGTTTTACCTTACTGGACCATCCCAATGCTACC
ACCGCATCAATCCCTCATCGCTATAAGCCAAAGTGTCAAATACAACTCTCCGACTCT
CGGAAAGCATCCGAGAGATGAAGGATCATCGAAAAAGGACAATGAGACAGAGCGAAAACA
GAAGGCCGGGTGCGTTCTGGTCCCGAAAACGTTGAGAATAGATGATCCTAACGAAGCAGC
AAAGAGCTCGATATGGACAACATTGGGAATCAAGAACGAGGCGATGTGCAAAGCCGGTGG
TATGTTCAAAGGGTTTGATCATAAGACAAAGATGTATAACAACGACAAAGCTGAGAACTC
CCCTGTTCTTTCTGCTAACCTGCTGCTCTATCAAGATCACACAATTTCCATGAACAGAT
TTAGAGTTACATATGTATATGTATATGTATGATTGATTGTATGTATAGATGATACTGG
AGAATGATGAGTTTTTGAAGTCAAACCTTTTCTTCTTTCTAGTGATTGCCTTTATTCC
TTTACATGTTTTGGTTCTCTGTACACTATTTGATTTACCTTTTTTACTTTCTTTCTTCAT
TTGTGAGGAAATGTTGGAAGATAACATTAATGGTAAAAAGTTGGTGTGGACCGTTGTGCG
GTTGGCATTTCAAAAAA

>G748 Amino Acid Sequence (domain in AA coordinates: 112-140)

MMMETRDPAILKPGMKIPFSPVFESAVTVEDDEDDWSGGDDKSPEKVTPELSDKNNNNC
NDNSFNNSKPEITLDKEEATSTDQIESDTPEDNQQTTPDGKTLKKPTKILPCPRCKSMET
KFCYYNNYINQPRHFCKACQRYWTAGGTMNRNVPVGAGRRKNKSSSSHYRHITISEALEA
ARLDPLGQANTRVLSFGLAQQHVAPMTPVMKLQEDQKVSNGARNRFHGLADQRLVAR
VENGGDCSSGSSVTTNNHSDVESRAQSGSVVEAQMNNNNNMNGYACIPGVWPYTW
PAMPPPGFYPPPGYPMFPYPTIPLPPLHQSSSPISQKCSNTNSPTLGKHPRDEGSSKK
DNETERKQKAGCVLPKTLRIDDPNEAAKSSIWTTLGIKNEAMCKAGGMFKGPDHKTMY
NNDKAENSPVLSANPAALSRSHNFHEQI*

>G247 (1..660)

ATGAGAATGACAAGAGATGGAAAAGAACATGAATACAAGAAAGGTTTATGGACAGTGGAA
GAAGACAAGATCCTCATGGATTATGTCCGAACTCATGGCCAGGGCCACTGGAACCGCATC
GCCAAGAAAACCTGGGCTCAAGAGATGTGGGAAAAGCTGTAGGTTGAGATGGATGAACCTAC
TTAAGCCCTAATGTTAAACAGAGGCAATTTTACTGACCAAGAAGAAGATCTCATCATCAGA
CTCCACAAGCTCCTCGGCAACAGATGGTCGTTGATAGCGAAAAGAGTTCCGGGAAGAACA
GACAACCAAGTAAAGAACTTACTGGAACACACATCTCAGCAAGAAACTTGGTCTCGGAGAT
CATTCAACTGCCGTCAAAGCCGCATGCGGTGTAGAGTCTCCACCGTCTATGGCCCTTATA
ACCACAACGTCCTCCTCTCATCAAGAGATCTCCGGTGGAAAAAATCAACTCTAAGGTTT
GACACTTTAGTTGACGAATCCAAACTCAAACCAAATCCAACTAGTCCACGCAACACCA
ACTGACGTAGAAGTTGCAGCTACGGTTCCAAATCTGTTTCGATAACCTTTGGGTTCTTGAA
GACGACTTCGAGCTTAGTTCACTCACTATGATGGATTTTACTAATGGGTATTGCCTTTGA
>G247 Amino Acid Sequence (domain in AA coordinates: 15-116)
MRMTRDGKEHEYKKGLWTVVEEDKILMDYVRTHGQGHWNRIAKKTGLKRCGKSCRLRWMNY
LSPNVNVRNFTDQEEELIIRLHKLLGNRWSLIAKRVPGRTDNQVKNYWNTHLSKKLGLGD
HSTAVKAACGVESPPSMALITTTSSSHQEISGGKNSLRFDTLVDESCLKPKSKLVHATP
TDVEVAATVPNLFDTFWVLEDDFELSSLTMMDFTNGYCL*

>G585 (111..2039)

CTCTCAAACATTTCTCTGTTTGTTCGGCGAAAACGGCAACTGTTTCATCAAATGACAAA
CACAAAAACCTTAACATCTAGTTTGTATCCTCTCTGATACTTCAAAAAAATGGATGAAG
AAACAATGGCTACCGGACAAAACAGAACAACTGTGCCAGAGAATCTGAAGAAACACCTCG
CAGTTTTCAGTTTCGAAACATTCATGGAGTTATGGTATCTTTTGGTCTGTCTCTGCTTCTC
AGTCTGGAGTTTTAGAATGGGGAGATGGATACTATAATGGAGATATCAAAACGAGGAAGA
CGATTCAAGCTTCGGAGATCAAAGCTGATCAGCTTGGTCTACGGAGGAGCGAGCAGCTTA
GCGAGCTTTACGAGTCTCTCTCCGTCGCTGAATCTTCTTCTTCAGGCGTTGCTGCCGGAT
CTCAAGTCACCAGACGAGCTTCCGCCGCCGCACTTTCACCGGAAGATCTCGCCGACACCG
AGTGGTACTATTTGGTTTGTATGTCTTTCGTCTTCAACATGGTGAAGGAATGCCTGGAC
GGACGTTTTCGAAACGGTGAACCGATATGGTTGTGCAACGCTCATACGGCGGATAGTAAAG
TGTTTAGCCGTTCTCTCTTAGCAAAAAGTGCTGCGGTTAAGACAGTGGTTTGTCTCCCGT
TCCTTGGAGGAGTCGTTGAGATTGGTACCACAGAACATATTACGGAAGACATGAATGTAA
TACAATGCGTGAAGACATCATTCCTCGAAGCCCCTGATCCGTACGCTACAATATTACCAG
CAAGATCCGATTATCACATCGACAACGTTCTTGATCCGCAACAGATTCTAGGCGACGAGA
TTTACGCGCCTATGTTTCAGTACGGAGCCTTTTCCAACAGCTTCTCCGAGCAGAACTACCA
ACGGTTTTCGATCAAGAACATGAACAAGTAGCAGATGATCATGATTCTTTCATGACCGAAA
GAATCACTGGAGGAGCTTCTCAGGTGCAAAGCTGGCAGCTCATGGACGACGAGCTTAGTA
ACTGCGTTACCAGTCTGCTAAATCCAGCGATTGCGTCTCTCAAACGTTTGTGAAGGGG
CGGCTGGACGGGTTGCTTACGGTGCAAGAAAGAGTAGAGTTCAAAGACTAGGGCAAATTC
AAGAGCAACAGAGAAATGTGAAGACATTGTCAATTTGATCCAAGAAACGACGACGTTTATT
ACCAAAGTGTGATCTCAACGATTTTAAAGACCAACCATCAGTTAATTCTCGGACCGCAGT
TTCGAAACTGCGATAAACAGTCAAGCTTCACTAGGTGGAAGAAATCATCGTCATCATCAT
CAGGAACCGCCACGGTCACGGCACCATCACAAGGAATGTTAAAGAAAATTATTTTCGATG
TTCCGCGAGTGCAACAGAGAGAGTAAATGTTGGACTCACCAGAAGCCAGAGATGAAA
CTGGGAACCATGCGGTTTTAGAGAAGAAGCGCCGAGAAATTGAACGAACGGTTTCATGA
CCTTGAGAAAAATCATTCCTCAATCAACAAGATCGATAAAGTATCGATTCTTGACGATA
CGATAGAGTATCTTCAAGAAGTCTGAGAGACGGGTTCAAGAAGTGAATCTTGACAGAGAT
CAACCGATACAGAGACTCGTGGGACGATGACGATGAAGAGGAAGAAACCATGCGACGCAG
GAGAAAGAACATCAGCTAATTGCGCAAATAATGAAACAGGAAATGGGAAGAAGGTGTCCG
TTAACAATGTTGGTGAAGCCGAGCCAGCAGATACCGGTTTTACTGGTTTAAACCGATAATT
TAAGGATCGGTTCCGTTTGGTAATGAGGTGGTTATTGAGCTTAGATGTGCTTGGAGAGAAG
GAGTATTGCTTGAGATAATGGATGTGATTAGTGATCTCCATTTGGATTCTCATTCCGTTT
AATCCTCGACCGGAGACGGTTTGTCTGCTTAACCGTCAATTGCAAGCACAGGGGTCAA
AAATAGCGACACCAGGAATGATCAAAGAAGCACTTCAAAGGGTTGCATGGATCTGTTGAA
GACTACTTAGTTAAAATTGACAGCAAAGAAAAAACATTCCCGGTTTGGTTTCTATTCTTT
GGTTTTCTTCTAACCGGGTTTTAGGAATTAATGTTATGTTTATCATTTGTTTTTTGTTTT
TTTTTTGTGCTTTTTTTCCGTTGCTTAACGTAGGTGAAGAGGAACATACACTATGCGTA
TTTTGTTTGAAGGTAGATTATTTAAGGGTATTAGTAATAGTAATAGCCAGTTTAGATGAT
TTTGTGTTCTTTTGTGTT

>G585 Amino Acid Sequence (domain in AA coordinates:436-501)
MDEETMATGQNRTTVPENLKKHLAVSVRNIIQWSYGIFWSVSASQSGVLEWGDGYNGDIK
TRKTIQASEIKADQLGLRRSEQLSELYESLSVAESSSSGVAAGSQVTRRASAAALSPEDL
ADTEWYYLVCMSFVFNIGEGMPGRTFANGEPIWLCNAHTADSKVFSRSLAKSAAVKTVV
CFPFLLGGVVEIGTTEHITEDMNVIQCVKTSFLEAPDPYATILPARSDYHIDNVLDPPQIL
GDEIYAPMFSTEPFPTASPSRTTNGFDQEHEQVADDHDSFMTERITGGASQVQSWQLMDD
ELSNCVHQSLNSSDCVSQTFVEGAAGRVAYGARKSRVQRLGQIQEQQRNVKTLSTFDPRND
DVHYQSVISTIFKTNHQLILGPQFRNCDKQSSFTRWKSSSSSSGTATVTAPSQGMLKKI
IFDVPRVHQKEKLMLDSPPEARDETGNHAVLEKKRREKLNERFMTLRKIIPSINKIDKVS
LDDTIEYLQELERRVQELESCRESTDTETRGTMTKRKKPCDAGERTSANCANNETGNGK
KVSVMNVGEAEPADTGFTGLTDNLRIGSFGNEVVIELRCAWREGVLLBIMDVISDLHLDS
HSVQSSTGDGLLCLTVNCKHKGSKIATPGMIKEALQRVAWIC*

>G634 (1..798)

ATGGAGCAAGGAGGAGGTGGTGGTGAATGAAGTTGTGGAGGAAGCTTCACTATTAGT
TCAAGACCTCCTGCTAACAACTTAGAAGAGCTTATGAGATTCTCAGCCGCCGCGGATGAC
GGTGGATTAGGAGGTGGAGGTGGAGGAGGAGGAGGAAGTGCTTCTTCTCATCGGGA
AATCGATGGCCGAGAGAAGAACTTTAGCTCTTCTTCTCGGATCCGATCCGATATGGATTCT
ACTTTTCGTGATGCTACTCTCAAAGCTCCTCTTGGGAACATGTTTCCAGGAAGCTATTG
GAGTTAGGTTACAAACGAAGTTCAAAGAAATGCAAAGAGAAATTGAAAACGTTTCAGAAA
TATTACAAACGTACTAAAGAACTCGCGGTGGTCGTCATGATGGTAAAGCTTACAAGTTC
TTCTCTCAGCTTGAAGCTCTCAACTACTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT
CATCAACCAGAACAGAAACAAACAACCAACCAACAAGAGATGGTCATGAGCTCGGAA
CAATCATCATTACCATCATCATCAAGATGGCCAAAGGCAGAGATTCTAGCGCTTATAAAC
CTGAGAAGTGAATGGAACCAAGGTACCAAGATAATGTACCTAAAGGACTTCTATGGGAA
GAGATCTCAACTTCAATGAAGAGAATGGGATACAACAGAAACGCTAAGAGATGTAAAGAG
AAATGGGAAAACATAAAACAAATACTACAAGAAAGTTAAAGAAAGCAACAACAGCAACTAC
ACAACAAGAATCAATGA

>G634 Amino Acid Sequence (domain in aa coordinates: 62-147, 189-245)

MEQGGGGGGNEVVVEEASPISSRPPANNLEELMRFSAAADDGGLGGGGGGGGGSSSSSG
NRWPREETLALLRIRSDMDSTFRDATLKAPLWEHVSRKLELGYKRSSKKCKEKFENVQK
YYKRTKETRGGRHDKAYKFFSQLEALNTTPPPPSHPHAHQPEQKQQQPQEQEMVMSSE
QSSLPSSSRWPKAEILALINLRSGMEPRYQDNVPKGLLWBEISTSMKRMGYNRNAKRCKE
KWENINKYKVKESNNSNYNNKNQ*

>G676 (1..612)

atgagaaagaaagtaagtagtagtggtgacgaaggaaacaatgagtacaagaaaggtttg
tggacagtagaagaagacaaaatcctcatggattatgtcaaagctcatggcaaggtcac
tggaatcgatttgcaaaaagactggtttaaagagatgtggaagagttgtagattgagg
tggatgaattatctcagccctaattgtgaaaagaggcaatttcaccgagcaagaagaggat
cttatcatttaggtcccaagttgcttggtaataggtggtctttaattgctaaaagagt
ccgggtcgaaacggataatcaagtgaagaactattggaacacgcatcttagtaagaaactc
ggaatcaaagatcagaaaaccaaacagagcaatggtgatattgtttatcaaataatctc
ccgaatcctaccgaaacatcagaagaaacgaaaatctcgaatattgtcgataacaataat
atcctcggagatgaaattcaagaagatcatcaaggaagtaactacttgagttcactttgg
gttcatgaggatgagtttgagcttagcacactcaccaacatgatggactttatagatgga
cactgtttttga

>G676 Amino Acid Sequence (domain in AA coordinates: 17-119)

MRKKVSSSGDEGNNEYKGLWTVVEEDKILMDYVKAHGKGHWNRIAKKTGLKRCGKSCRLR
WMNYLSPNVKRGNFTEQEEDLIIRLHKLLGNRWSLIAKRVPGRTDNQVKNYWNTHLSKKL
GIKDQKTKQSNQDIVYQINLPNPTETSEETKISNIVDNNILGDEIQEDHQGSNYLSSLW
VHEDEFELSTLTNMMDFIDGHCF*

>G682 (1..228)

ATGGATAACCATCGCAGGACTAAGCAACCCAAGACCAACTCCATCGTTACTTCTTCTTCT
GAAGAAGTGAGTAGTCTTGAGTGGGAAGTTGTGAACATGAGTCAAGAAGAAGAAGATTG
GTCTCTCGAATGCATAAGCTTGTCTGGTGACAGGTGGGAACGATAGCTGGGAGGATCCCA
GGAAGAACCCTGGAGAAATTGAGAGGTTTTGGGTGATGAAAAATTGA

>G682 Amino Acid Sequence (domain in AA coordinates 27-63)

MDNHRRTKQPKTNSIVTSSEEVSSLEWVNVMSQEEEDLVSRMHKLVDGRWELIAGRIP

GRTAGEIERFVWMKN*

>G635 (1..993)

ATGGAGATCATGCGTCCAGGGGTCTCAGAAAACACTTTGAAAGGAAAAATAAGAATCACA
ACGCGGTGCATGTGGCTTGACAAAGGAAGACTTTTAGATGCACCTTCACAAAGCAGCTCAT
GCTGCTCTATCAAGTTGTCTGTGACATGTCCCTTGCTCAGATGGAAAGAACAGTCTCC
GAAGTCTTGAGGAAGATTGTAAGGAAGTACAGTGGTAAAAGGCCTGAAGTCATCGCTATA
GCCACTGAGAATCCAATGGCTGTCCGAGCTGATGAGGTCAGTGCAGACTGTCTGGTGAT
CCAAGTGTGGTTCTGAGATTGCAGCTTTAAGGAAAGTTGTTGAAGGAAATGACAAAAGA
AGTCGGGCGAAGAAAGCACCTTCACAAGAAGCTTCCCCCAAAGAAGTAGATCGCACTTTG
GAAGATGATATCATTGATAGTGCAAGACTACTGGCTGAAGAAGAACTGCGGCATCAACA
TACACGGAAGAAGTTGATACGCCCGTTGGGAGTTCTTCAGAAGAGTCAGACGATTTTGG
AAATCATTTCATCAATCCATCATCGTCACCTTCACCGAGTGAAACAGAAAATATGAATAAG
GTAGCTGATACGGAGCCTAAAGCAGAGGGTAAGGAAAACAGCAGAGACGACGATGAATTA
GCTGATGCTTCAGATTCTGAAACCAAGTCATCACCAAAACGTGTGAGGAAGAACAATGG
AAACCGGAGGAGATAAAGAAGGTAATCAGAATGCGAGGAGAGCTGCACAGTAGATTTCAA
GTGGTGAAAGGTAGAATGGCATTGTGGAAGAGATCTCTCAAATCTATCAGCTGAAGGA
ATCAATCGAAGCCCGGGACAATGCAAATCTCTCTGGGCATCACTTATTCAGAAATACGAG
GAGAGCAAGGCTGATGAGAGAAGCAAGACGAGTTGGCCACATTTTGAGGATATGAACAAC
ATTTTGTCTAGAGCTAGGCACACCTGCGTCTTAA

>G635 Amino Acid Sequence (domain in AA coordinates: 239-323)
MEIMRPGVSENTLKGKIRITTRCMWLDKGRLLDALHKAHAALSSCPVTCPLSHMERTVS
EVLRKIVRKYSGRPEVIAIATENPMAVRADSVSARLSGDPVSGVAAALRKVVEGNDKR
SRAKKAPSQEASPKVDRTLEDDIIDSARLLAEETAASTYTEEVDTPVGSSEESDDFW
KSFINPSSSPSETENMNKVADTEPKAEGKENSRRDDELADASDSETKSSPKVRKNKW
KPKEIKKVIRMRGELHSRFQVVKGRMALWEEISSNLSAEGINRSPGQCKSLWASLIQKYE
ESKADERSKTSWPHFEDMNNILSELGTPAS*

>G1068 (150..1310)

GAGAGTTGTTAGCTAGCTACACGCTTTCGCTTAAACTCAAAAACCTGCACTTTCTCGT
CTATTTTCTCGGCATTTCGTAAAACAGAAAAGTGGGTCTCCAAGAAAATTACCTTAAATTC
ACAAAGATTTCATCTTTTCTCCACCTCCAATGGATTCCAGAGAGATCCACCACCAACAAC
AGCAACAACAACAACAACAACAGCAGCAGCAGCAACAACAGCAACATCTACAACAACAGC
AACAACCACCGCCAGGGATGTTAATGAGTCACCACAATTCCTACAATCGAAACCTTAACG
CCGCCGCCGCTGTTTAAATGGGTACAAACACCTCCACATCTCAAGCTATGCATCAAAGAT
TACCTTTTGGTGGTTCTATGTACCCGCATCAGCCTCAACAACATCAGTATCATCATCCTC
AGCCTCAGCAACAGATAGATCAGAAGACTCTTGAATCTCTTGGATTTCCTACTTCGCCTC
TTCCTTCTGCTTCTAATCTTACGGTGGTGAAATGAAGGAGGTGGTGGTGGTGTAGCG
CCGGAGCTAATGCTAATCTTCCGATCCACCTGCTAAACGGAACAGAGGACGTCCTCCTG
GCTCCGGTAAGAAGCAGCTCGATGCTTTAGGAGGAACAGGAGGAGTTGGGTTCACGCCTC
ATGTCAATTGAGGTTAAAACAGGAGAGGACATAGCTACGAAGATATTGGCGTTTACGAACC
AAGGGCCACGCGCAATCTGTATTCTCTCAGCTACAGGAGCTGTAATAATGTGATGCTTC
GTCAAGCTAACAATAGCAATCCTACTGGAAGTGTAAAGTATGAGGGCCGATTGAAATCA
TTTCTCTGTCTAGGTTCTTTCTTGAATTCTGAGAGTAATGGTACTGTGACCAAACTGGTA
ACTTGAGTGTGTCGCTGGCTGGACACGAAGGCCGATTGTGGGTGGATGTGTTGATGGAA
TGCTAGTAGCTGGATCACAAGTCCAGGTCATTGTGGGAAGCTTTGTACCAGATGGAAGGA
AGCAGAAACAAGTGCGGGGCGTGCTCAGAATACTCCGGAGCCAGCTTCAGCACCAGCCA
ATATGTTGAGCTTTGGTGGTGGTGGTGGACCGGAAGCCCTCGATCTCAAGGACAACAAC
ACTCGAGCGAGTCATCAGAGGAAAACGAAAGTAATTCTCCGTTGCACCGTAGAAGCAACA
ACAACAACAGCAACAATCATGGGATATTTGGAACTCTACACCTCAACCGCTTCACCAA
TTCCTATGCAGATGTACCAGAATCTCTGGCTGGCAACAGTCTCAATAAACAGATGGTT
CATGGGTCAAGATTTGACCGGGTTTGCTTCTGTCTCTTTTGACACATCTCTCCATCAG
ATTTATCTCTATAAAGTAGATTGAGCTCTTACTCTCTCATCTCTCTCTCTTTACTAT
TTCTCTTAAATTTAGCTTTGGTTTTAGATAAATAGAGAGAGAGACATGTTAAGTAGGT
TTCAAATTCATCTTGTGTTTGTGTTTCTTAGTAGTTTCTTTTGATTGTGATGATCATA
AAGACTTGTTCTTTTCTCTATATTCAACGAATTATCCACTTTAA

>G1068 Amino Acid Sequence (domain in AA coordinates: 143-150)
MDSREIHHQQQQQQQQQQQQQQQHLQQQQPPPPGMLMSHHNSYNRNPNAAAVLMGHN
TSTSQAMHQRLPFGGSMSPHQPQHQYHHPQPPQQIDQKTLES LGFP TSPLPSASNSYGG

GNEGGGGGDSAGANANSSDPPAKRNRGRPPGSGKKQLDALGGTGGVGFTPHVIEVKTGED
IATKILAFNTNQGPRAICILSATGAVTNVMLRQANNSNPTGTVKYEGRFEIISLSGSFLNS
ESNGTIVTKTGNLSVSLAGHEGRIVGGCVDMVLVAGSQVQVIVGSFVDPGRKQKQSAGRAQ
NTPEPASAPANMLSFGGVGGPGSPRSQGGQHSSESSEENESNSPLHRRSNNNNNSNNHGIF
GNSTPQPLHQIPMQMYQNLWPGNSPQ*

>G1225 (1..984)

ATGACTCTAGAAGCTTTATCATCAAACGGTCTTTTAAACTTTTGTCTCTGAAACTCTT
TCACCAACTCCATTCAAGTCTCTCGTCGATCTCGAGCCATTGCCGAAAATGATGTCATC
ATATCGAAGAACACAATTTTCGGAGATATCTAATCAAGAACCGCCACCACAGCGACAACCA
CCAGCTACGAATCGAGGGAAGAAGCGGCGGAGGAGGAAGCCTAGGGTTTGCAAAAACGAG
GAAGAAGCTGAGAATCAACGAATGACTCACATTGCCGTCGAAAGAAATCGAAGAAGACAA
ATGAATCAACATCTCTCTGTCTTGCATCTCTCATGCCTCAACCTTTTGTCTCACAAGGGT
GATCAAGCTTCAATAGTTGGTGGAGCCATAGATTTCATCAAAGAAGCTTGAACACAAATTA
CTATCTCTTGAAGCTCAAAAACATAGATAATGCTAAATTAACCAGTCGGTTACTTCTTCA
ACAAGTCAAGACTCAATGGTGAACAAGAGAATCCTCATCAACCATCTTCACTATCTCTA
TCGCAGTTCTTTCTTCAATCATACGATCCGAGCCAAGAGAATAGGAACGGCTCAACAAGC
TCGGTGAAAACCCCTATGGAAGATCTTGAGGTGACTCTAATCGAAACTCATGCTAACATC
AGAATCTTGTCGAGAAGAGAGGTTTCCGGTGGAGCACGTTGGCCACCACCAAACCGCCG
CAGCTTTCGAAGCTGGTGGCTTCTCTACAATCGCTGTCCCTCTCCATTCTTCACCTTAGT
GTCACAACATTGGACAATTATGCTATTTACTCCATCAGCGCTAAGGTGGAAGAGAGTTGC
CAGCTAAGTTTCAGTAGATGACATTGCGAGGAGCAGTTACCACATGCTAAGTATCATTGAA
GAGGAGCCTTTTGTGTTGCTCATCAATGTGAGAATTACCATTGACTTCTCTTTGAATCAC
TCAATGTCACTCATCTCTCTGAGAAATCTCTTTTGTGTTGTTATTCCTTCTTTTA
ATTTTATCACATAGCACATCTTTAGTTTTTTTTTTTT

>G1225 Amino Acid Sequence (domain in AA coordinates: 78-147)

MTLEALSSNGLLNFLSPTLSPFVKSLVDLEPLPENDVIISKNTISEISNQEPQPQRQP
PATNRGKKRRRRKPRVKNEEAENQRMTHIAVERNRRRQMNQHLVLRSLMPQPFHAKG
DQASIVGGAIIDFIKELEHKLSSLEAQKHNAKLNSVTSSTSQDSNGEQENPHQPSLSL
SQFFLHSYDPSQENRNGSTSSVKTPMEDLEVTLIETHANIRILSRRRGFRWSTLATTKPP
QLSKLVASLQSLSLHLVSVTLTDNYAIYSISAKVEESCQLSSVDDIAGAVHMLSIIE
EFPFCCSSMSELPDFSLNHSNVTHSL*

>G1337 (97..1398)

AATGGATTGTGCATCATCTCTCACCGTCCTTAGTCTCTGAAAATAAATCTGATTTTG
ATTTGCAATTTAGGGATTTTGAAGAGAGTCAATTATGAGTAGTTCGGAGAGAGTACCG
TGCGATTTCTGCGGCGAGCGTACGGCGGTTTTGTTTTGTAGAGCCGATACGGCGAAGCTG
TGTTTGCCCTGTGATCAGCAAGTTCACACGGCGAATCTGTTGTCGAGGAAGCACGTGCGA
TCTCAGATCTGCGATAATTGCGGTAACGAGCCAGTCTCTGTTGCGGTGTTTACCGGATAAT
CTGATTTTGTGTCAGGAGTGTGATTGGGATGTTACGGAAGTTGTTCACTTTCCGATGCT
CATGTTTCGATCCGCCGTGGAAGGTTTTTCCGGTGTCCATCGGCGTTGGAGCTTGCTGCT
TTATGGGGACTTGATTGGAGCAAGGGAGGAAAGATGAAGAGAATCAAGTTCCGATGATG
GCGATGATGATGGATAATTTCCGGATGCAGTTGGATTCTTGGGTTTTGGGATCTAATGAA
TTGATTGTTCCAGCGATACGACGTTTAAGAAGCGTGGATCTTGTGGATCTAGTTGTGGG
AGGTATAAGCAGGTATTGTGTAAGCAGCTTGAGGAGTTGCTTAAGAGTGGTGTGTGCGGT
GGTGATGGCGATGATGGTGATCGTGACCGTGATTGTGACCGTGAGGGTGCTTGTGATGGA
GATGGAGATGGAGAAGCAGGAGAGGGGCTTATGGTTCCGAGATGTCAGAGAGATTGAAA
TGGTCAAGAGATGTTGAGGAGATCAATGGTGGCGGAGGAGGAGTTAACCAGCAGTGG
AATGCTACTACTACTAATCCTAGTGGTGGCCAGAGTTCTCAGATATGGGATTTTAACTTG
GGACAGTCACGGGGACCTGAGGATACGAGTCGAGTGGAAGCTGCATATGTAGGGAAAGGT
GCTGCTTCTTCAATCACAATCAACAATTTGTTGACCATATGAATGAAACTTGTTCCTACT
AATGTGAAAGGTGTCAAAGAGATTAAAAAGGATGACTACAAGCGATCAACTTCAGGCCAG
GTACAACCAACAAAATCTGAGAGCAACAATCGTCCAATFACCTTTGGCTCTGAGAAAGGT
TCGAACTCCTCCAGTGACTTGCAATTCACAGAGCATATTGCTGGAAC TAGTTGTAAGACC
ACAAGACTAGTTGCAACTAAGGCTGATCTGGAGCGGCTGGCTCAGAACAGAGGAGATGCA
ATGACAGCTTACAAGGAAAAGAGGAAGACACGAGATATGATAAGACCATAAGGTATGAA
TCGAGGAAGGCAAGAGCTGACACTAGGTTGCGTGTGAGGCGAGATTGTGAAAGCTAGT
GAAGCTCCTTACCCTTAACCTTAAGTTTTTTTACATAGGCTTCTTTTAGCTACAAACTT
AGTTACTTTTTTTTACTCCACTGCCTCATAAATGTACAGACCGGTCTCGTTTCATCTGGCC

GCCCTTCTTGTGTTTTATTGCCTTATCTGGCCCTTTTATGTACCTTGAATCTTATCTAGTT
TAAAAAAGATTGTAACCTTCTAGAAAACCATATTCTGTTGACAGTATATACATGTCTATC
CAAGCAAAAA

>G1337 Amino Acid Sequence (domain in AA coordinates: 9-75)
MSSSERVPCDFCGERTAVLFCRADTAKLCLPCDQVHTANLLSRKHVRSQICDNCNEPV
SVRCFTDNLILCQECDWDVHGSCSVSDAHVRSVEGFGCPSALELAALWGLDLEQGRKD
EENQVPMAMMDNFGMLDSWVLGSNELIVPSDTTFKKRGSCGSSCGRYKQVLCKQLEE
LLKSGVVGGDGGDDGDRDRCDREGACDGDGDGEAGEGLMVPMSERLKWSRDVEEINGGG
GGGVNQQWNATTTNPSGGQSSQIWDNFLGQSRGPEDTSRVEAAYVGKGAASSFTINNFD
HMNETCSTNVKGVKEIKDDYKRSTSGVQPTKSESNNRPITFGSEKGSNSSSDLHFTTEH
IAGTSCKTTRLVATKADLERLAQNRGDAMQRYKEKRKTRRYDKTIRYESRKARADTRLRV
RGRFVKASEAPYP*

>G1759 (110..700)

CGAGAAAAGGAAAAAATAAGAAAGAGAAAACGCTTAGTATCTCCGGCGACTTGAAC
CCAAACCTGAGGATCAAATTAGGGCACAAAGCCCTCTCGGAGAGAAGCCATGGGAAGAAA
AAACTAGAAATCAAGCGAATTGAGAACAAAAGTAGCCGACAAGTCACCTTCTCCAAACG
TCGCAACGGTCTCATCGAGAAAGCTCGTCAGCTTCTGTTCTCTGTGACGCATCCGTCGC
TCTTCTCGTCTCTCCGCCTCCGGCAAGCTCTACAGCTTCTCCTCCGGCGATAACCTGGT
CAAGATCCTTGATCGATATGGGAAACAGCATGCTGATGATCTTAAAGCCTTGGATCATCA
GTCAAAAGCTCTGAACCTATGGTTCACACTATGAGCTACTTGAACCTTGGATAGCAAGCT
TGTGGGATCAAATGTCAAAATGTGAGTATCGATGCTCTTGTTCAACTGGAGGAACACCT
TGAGACTGCCCTCTCCGTGACTAGAGCCAAGAACCGAAGCTCATGTTGAAGCTTGTGGA
GAATCTTAAAGAAAAGGAGAAAATGCTGAAAGAAAGAGAACCAGGTTTGGCTAGCCAGAT
GGAGAATAATCATCATGTGGGAGCAGAAGCTGAGATGGAGATGTCACCTGCTGGACAAAT
CTCCGACAATCTTCCGGTGACTCTCCCACTACTTAATTAGCCACCTTAAATCGGCGGTTG
AAATCAAAATCCAAACATATATAATTATGAAGAAAAAATAAGATATGTAATTATT
CCGCTGATAAGGGCGAGCGTTTGTATATCTTAATACTCTCTCTTTGGCCAAGAGACTTTG
TGTGTGATACTTAAGTAGACGGAAGTAAGTCAATACTATCTGTTTTAAGACAAAAGGTTG
ATGAACCTTGTACCTTATTCGTGTGAGAAAAA

>G1759 Amino Acid Sequence (conserved domain in AA coordinates: 2-57)

MGRKLEIKRIENKSSRQVTFKRRNGLIEKARQLSVLCDASVALLVVSASGKLYSFSSG
DNLVKILDYRGKQHADDLKALDHQSKALNYGSHYELLELVDSKLVGSNVKNVSI DALVQL
EEHLETALSVTRAKTEMLKLVENLKEKEKMLKEENQVLASQMNHHVGAEEEMESP
AGQISDNLPLPLN*

>G1804 (169..1497)

TATCTCTCTCTTTCTCAAAACCTTTTCAGTCAAAATCTCTCCGGCGGCTTTTAAACTATGTG
AAGGAGGAGAACCTCCATAACAAGAAGCGGATTCTCTCAGTTTTCCGGCGGCGAGGAAC
ACAAAGCCACCGGTTTGTAGACACAGATTTTCATTTTCAGTTGTTAAATGGTAAGTAGA
GAAACGAAAGTTGAGTTCAGAGCGAGAAGTAGAGTCCATGGCGCAAGCGAGACATAAT
GGAGGAGGTGGTGGTGAGAATCATCCGTTTACTTCTTTGGGAAGACAATCCTCTATCTAC
TCATTGACCTTGACGAGTTCCAACATGCTTTATGTGAGAACGGCAAGAACTTTGGGTCC
ATGAACATGGACGAGTTTCTTGTCTCTATTTGGAACGCAGAGGAGAATAATAACAATCAA
CAACAAGCAGCAGCAGCTGCAGGTTTCAATCTGTTCCGGCTAATCACATGGTTC AAC
AACAACAATAACAATGGAGGCGAGGGTGGTGTGGTGTCTTTAGTGGTGGTCTAGAGGC
AACGAAGATGCTAACAATAAGAGAGGGATAGCGAACGAGTCTAGTCTTCCTCGACAAGGC
TCTTTGACACTTCCAGCTCCGCTTTGTAGGAAGACTGTTGATGAGGTTTGGTCTGAGATA
CATAGAGGTGGTGGTAGCGGTAATGGAGGAGACAGCAATGGACGTAGTAGTAGTAAT
GGACAGAACAAATGCTCAGAACGGCGGTGAGACTGCGGCTAGACAACCGACTTTTGGAGAG
ATGACACTTGAGGATTTCTTGGTGAAGGCTGGTGTGGTGTAGAGAACATCCCACTAATCCT
AAACCTAATCCAAACCCGAACCAAAACCAAAACCGTCTAGTGTAATACCCGCAGCTGCA
CAGCAACAGCTTTATGGTGTGTTTCAAGGAACCGGTGATCCTTCATTCCTGGGTCAAGCT
ATGGGTGTGGGTGACCCATCAGGTTATGCTAAAAGGACAGGAGGAGGAGGTATCAGCAG
GCGCCACCAAGTTCAGGCAGGTGTTGCTATGGAGGTGGCGTTGGGTTTGGAGCGGGTGA
CAGCAATGGGAATGGTTGGACCGTTAAGCCCGGTGCTTTCAGATGGATTAGGACATGGA
CAAGTGGATAACATAGGAGGTGATGAGTAGATATGGAGAGGCTAAGGGGAAGGAAA
AGAGTAGTGGATGGTCCAGTGGAGAAAGTAGTGGAGAGAAGACAGAGGAGGATGATCAAG
AACC CGAGTCTGCTGCTAGATCTAGAGCAAGAAAACAAGCATATACAGTGAATTGGAA

GCTGAACCTTAACCAGTTGAAAGAAGAGAATGCGCAGCTAAAACATGCATTGGCGGAGTTG
GAGAGGAAGAGGAAGCAACAGTATTTTGTAGAGTTTGAAGTCAAGGGCACAACCGAAATTG
CCGAAATCGAACGGGAGATTGCGGACATTGATGAGGAACCCGAGTTGTCCACTCTAAACA
AACAAATAGGAAGATGGAGAAGAAGTCGGAGACAGAACGAGGAAAACTGATGATTTTCT
ACGTTGTTGTTTTGTCTTTGAGGAATGAGGTTATAGAATCTTTTACTTTGATGTTTTCT
GTGTTGGTAGGAGGAACACCATCTGATCTGCTTTACTAGTGTTCCTGTGAACAAAGAAA
GTGATTCTGTGTTTTCAACATCATCAATCTTTGGAAA

>G1804 Amino Acid Sequence (domain in AA coordinates: 357-407)

MVTRETKLTSEREVESSMAQARHNGGGGGENHPFTSLGRQSSIYSLTLDEFQHALCENGK
NFGSMNMDEFLVSIWNABENNNNQQAAGSHSVANHNGFNMMNNNGEGGVGVFSG
GSRGNEDANNKRGIANESSLPQGSLLTPAPLCRKTVDEVWSEIHRGGGSGNGGDSNGRS
SSSNGQNNAQNGGETAARQPTFGEMTLEDFLVKAGVVREHPTNPKNPNPNQNPSSVI
PAAQQQLYGVFQGTGDPSPFGQAMGVGDPSGYAKRTGGGGYQQAPPVQAGVCYGGGVGF
GAGGQQMGMVGPLSPVSSDGLGHGQVDNIGGQYGVDMGGLRGRKRVDGPVEKVVERRQR
RMIKNRESAARSARKQAYTVELEAELNQLKEENAQLKHALAELEKRRKQYFESLKSRA
QPKLPKSNGLRLTLMRNPSCPL*

>G207 (16..930)

aaaagatctgtttcaatggcggatcggtgtaaagggtccatggagtcagaagaagatgag
cagctacgaaggatgggttagaaatacggaccgaggaattgggtctcgattagcaaatcg
attccaggtcgatctggtaaatcggttagattacgttggtgtaatcagttatctccggag
gttgagcatcgctcttctcgccggaggaagatgagactattgtaaccgcccgtgctcag
tttggttaacaagtggcgacgattgctcgctcttcttaacggtcgtacggataacgccgtt
aaaaatcactcggactctacgcttaagaggaaatgcagcggaggtgtggcggttacgacg
gtgacggagacggaggaagatcaggatcgccgaagaaggagatctgttagctttgat
cctgcttttgctccggtggatactggattgtacatgagtcctgagagtcctaaccggaatc
gatgttagtgattctagcacgattccgtcaccgctgctcctggtgctcagctgtttaa
ccaatgccgatttccggcggttttacgggtggttccgcagccggttaccggttgaaatgtct
tcgctcttcggaggatccacctaacttcgttgagttgtcactacctggagctgagaacacg
agttcgagccataacaataacaacaacgcggttgatgtttccgagatttgagagtcagatg
aagattaatgtagaggagagaggaggaggagaaggacgtagaggtgagtttatgacg
gtgggtgcaggagatgataaaagctgaagtgaggagttacatggcggaaatgcagaaaaaca
agtgggtgatttcgctcgctcgagggtttatacgaatccggcggaatggtggttttagggat
tgtggagtaataacacctaaggttgagtagttttggttttaggggttaaaacttgaatcgat
tggggattttcaagagcattcatttttgggggtttatggtaaaattaaaaacaaaaacaaa
atgtacagaggaattaaaatttctatggaataatcttaaatctcaaatatttgttacttg
ttttggtgattcataacaaaaatcaaa

>G207 Amino Acid Sequence (domain in AA coordinates: 6-106)

MADRVKGPWSQEEDQLRRMVEKYGPRNWSAISKSIPIGRSGKSCRLRCNQLSPEVEHRP
FSPEEDETIVTARAQFGNKWATIIARLLNGRTDNAVKNHWNSTLKRKCSGGVAVTTVTETE
EDQDRPKRRRSVSFDPAFAPVDTGLYMSPESPNGIDVSDSSTIPSPSSPVAQLFKPMPIS
GGFTVVPQPLPVEMSSSSSEDPTSLSLSLPGAENTSSSHNNNNNALMFPRFESQMKNIVE
ERGGGGEGRRGEFMTVVQEMIKAEVRSYMAEMQKTSGGFVVGGGLYESGGNGGFRDCGVIT
PKVE*

>G218 (1..1182)

ATGGAGGCAGAGATCGTGAGACGATCGGAGGTAACGGGATTAAGAAGGGAGGTGGAAGAA
TCGTCATATTGGTAGAGGAGATTGCGATGGTGATGGCGGCGATGTGGGAGAAGATGCGGCA
GGGTTTCGTTGGGACGAGCGGGAGAGGAAGAAGAGATCGAGTTAAAGGGCCGTGGTCCAAG
GAGGAGGATGATGTGTTGAGTGAGCTCGTTAAGAGGTTGGGAGCGAGGAATTGGAGTTTT
ATCGCTCGGAGTATTCCTGGTCGTTCAAGCAAGTCTTGTCTCTCGTTGGTGTAATCAG
CTCAATCCAAATCTTATACGCAATTCATTTACTGAGGTAGAGGATCAGGCTATCATCGCA
GCACATGCCATCCACGGAAACAAATGGGCTGTTATCGCGAAGCTCCTCCCCGGAAGAACA
GATAATGCTATCAAGAACCACTGGAACCTCTGCTTTAAGACGTCGATTATAGACTTTGAA
AAGGCCAAGAATATAGGAACCTGGAAGCTTGCTCGTGATGATTCTGGATTTGACAGAACG
ACAACAGTAGCCTCATCAGAAGAACTTTATCTTCAGGCGGTGGTTGCCATGTAACACT
CCAATTGTATCTCCAGAAGGCAAAGAAGCTACCACCTCCATGGAAATGTCTGAAGAACAA
TGCGTAGAGAAAACAAACGGAGAAGGTATTTCTAGGCAAGATGATAAGGATCCTCCAACG
CTTTTCGCCCCAGTGCCTCGGCTCAGTCTTTTAAATGCTTGCAATCACATGGAAGGATCA

CCCTCTCCACATATACAAGACCAAAATCAGCTCCAATCATCTAAACAAGACGCAGCAATG
CTAAGATTGCTTGAAGGAGCTTACAGCGAACGGTTTGTGCCTCAAACATGTGGAGGTGGT
TGTTGCAGCAACAATCCCGATGGCAGTTTTCAGCAAGAATCATTGTTGGGTCCAGAGTTT
GTGGATTACTTAGACTCACCAACGTTTCCGAGTTCCGAAGTACGTGCTATAGCAACGGAA
ATAGGCAGCTCGCTTGGCTGAGAAGCGGTTTAGAGAGTAGCAGCGTGAGGGTGATGGAA
GACGCAGTTGGTCGGTTAAGGCCTCAAGGCTCCAGGGGTCATCGAGATCATTATCTTGTA
TCTGAACAGGGGACGAACATAACCAATGTCCTGTCCACATAA

>G218 Amino Acid Sequence (domain in AA coordinates: TBD)
MEAEIVRRSEVTGLRREVEESSIGRGDCDGDGGDVGEDAAGFVGTSGRGRRDRVKGPWSK
EEDDLVSELVKRLGARNWSFIARSIPGRSGKSCRLRWCNQLNPNLIRNSFTEVEDQAIIA
AHAIHGNKWAVIAKLLPGRTDNAIKHNWNSALRRRFIDFEKAKNIGTGLVVDSDSGFDRT
TTVASSEETLSSGGGCHVTTPIVSPGKEATTSMEMSEEQVEKTNGEGISRQDDKDPPT
LFRPVPRLSSFNACNHMEGSPSPHIQDQNLQSSKQDAAMLRLLEGAYSERFVPQTCGGG
CCSNPDGSGFQGESLLGPEFVDYLDSPTFPSSSELAIAIEIGSLAWLRSGLESSSVRVM
DAVGRLRPQSGRGRDHYLVSEQGTNITNVLST*

>G241 (46..867)
GAAAAAÇATTTCAACTTCTTTTATCAGCAATCACAAATCAAAGAGATGGGAAGAGCTCCA
TGCTGTGAGAGATGGGGTTGAAGAGAGGACCATGGACACCTGAAGAAGATCAAATCTTG
GTCTCTTTTATCCTCAACCATGGACATAGTAAGTGGCGAGCCCTCCCTAAGCAAGCTGGT
CTTTTGAGATGTGGAAGAAAGCTGTAGACTTAGGTGGATGAAGTATTTAAAGCCTGATATT
AAACGTGGCAATTTACCAAAGAAGAGGAAGATGCTATCATCAGCTTACACCAAATACTT
GGCAATAGATGGTCAGCGATTGCAGCAAAACTGCCTGGAAGAACCGATAACGAGATCAAG
AACGTATGGCACACTCATTGAAGAAGAGACTCGAAGATTATCAACCAGCTAAACCTAAG
ACCAGCAACAAAAAGAAGGGTACTAAACCAAATCTGAATCCGTAATAACGAGCTCGAAC
AGTACTAGAAGCGAATCGGAGCTAGCAGATTCATCAAACCTTCTGGAGAAAGCTTATTT
TCGACATCGCCTTCGACAAGTGAGGTTCTTCGATGACACTCATAAGCCACGACGGCTAT
AGCAACGAGATTAATATGGATAACAAACCGGGAGATATCAGTACTATCGATCAAGAATGT
GTTTCTTTTCGAACTTTTGGTGGGATATCGATGAAAGCTTCTGGAAGAGACTGTAT
AGCCAAGATGAACAACTACGTATCGAATGACCTAGAAGTCGCTGGTTTAGTTGAGATA
CAACAAGAGTTTCAAACTTGGGCTCCGCTAATAATGAGATGATTTTTGACAGTGAGATG
GAACCTTCTGGTTCGATGTATTGGCTAGAACCAGCGGGGAACAAGATCTCTTAGCCGGGCT
CTAGTTAACATGTTTGAGGAGTAAAGTGAAATGGTGCAAATAGTTAAGGCTAAGAAATT
CAAAAGCTTTTGTTTACCGAGAAAAAACACACTCTAACTCTTGATGTGATGTAGTTAGT
GTATTAATTAGAGGCTGCGTTTTCAA

>G241 Amino Acid Sequence (domain in AA coordinates: 14-114)
MGRAPCCEKMGKRGPTWPEEDQILVSFILNHGHSNWRALPKQAGLLRCGKSCRLRWMNY
LKPDKIKRGNFTKEEEDAIISLHQILGNRWSAIAAKLPGRTDNEIKNVWHTLKKRLEDYQ
PAKPKITSNKKGTGPKSESIVITSSNSTRSESELADSSNPSSGESLSTSPSTSEVSSMTLI
SHDGYSNEINMDNPKPGDISTIDQECVSFETFGADIDESFWKETLYSQDEHNYVSNLDLEVA
GLVEIQQEFQNLGSANNEMIFDSEMEILLVRCIG*

>G254 (15..923)
CGATTTTCAGCTCTATGGTGTCCGTAAACCTAGACCTAAGGGTTTTCCAGTTTTTCGATT
CCTCGAATATGAGTTTACCAAGCTCCGATGGATTGTTGGTTCGATTCCGGCCACGGGACGGA
CCAGTACGGTGTGTTTTCTGAGGATCCGACGACGAAGATTCGGAAGCCGTACACAATCA
AGAAGTCGAGAGAGAATTGGACAGATCAAGAGCACGATAAATTTCTAGAAGCTCTTCACT
TATTTCGATAGGGATTGGAAGAAAATAGAAGCCTTTGTTGGATCAAAAACAGTAGTTTCA
TACGAAGCCACGCTCAGAAATACCTTCTCAAAGTTCAGAAGAGTGGTGCTAACGAACATC
TTCCACTTCTCGACCTAAGAGGAAAGCGAGTCATCCTTATCCTATAAAGGCTCCTAAAA
ATGTTGCTTATACCTCTCTCCCGTCTTCGAGTACATTACCGTTGCTTGAGCCTGGTTATT
TGTATAGCTCTGATTGCAAGTCATTGATGGGAAACAGGCTGTTTGTGCATCTACCTCTT
CTTCGTGGAATCATGAATCGACAAATCTGCCAAAACCGGTGATTGAAGAGGAACCGGGAG
TCTCGGCCACGGCTCCTCTCCCAAATAATCGCTGCAGACAGGAAGATACAGAGAGGGTAC
GAGCAGTGACAAAGCCAAATAACGAAGAAAGTTGTGAAAAGCCACATAGAGTGATGCCGA
ATTTGCTGAAGTTTACAGCTTCAATTGGAAGTGTCTTCGATCCCAACACATCAGGCCACC
TCCAGAGATTAAGCAGATGGATCCAATAAATATGGAACGGTCTTTTTACTGATGCAAA
ACCTGTCTGTAAATCTGACAAGTCCCGAGTTTGAGAGCAAAGGAGTTGATATCATCAT
ACAGCGCTAAAGCTTTGAAATAGAGATAGAATAAAACAATAATGTACCTTATGTGAGATC

AAGAGACAATCATCCAAGGTCTGTATGCATTGCTTGGATTTAGGCCTCGTGTCTCACTA
CAGGAGCAGAACCAATCGCAAAGACTCTTAGATGGCTACTGAGTTGTGGTTTTATGTCT
CTGTAAGTCGCGGTGGAGCACACGTGTTTGTCTGTCTTGTGTATGTGTGTATAGATAAT
ACAAGGTTTTGCAGAGTAAGGTCACAGTTAGCTGCAAGTGAGTTTGGATCAATCTTAAGA
TTAAACCCCTGAGAGTGAGTGTCCTCAAAGAGACTGTGTAATATTGGTTTGGCGGTCAGCAG
AAGAGTTTTTGAAGTGACATCCAGTTAGTGATAACACGGTTGAAGAAAAGGTAAGGTTAC
AAGTTTAGTTTTGAATAATTGTATACTCAAAAAATATGAATGTATAAAGAATAATCACTT
GAGTCGCCTTA

>G254 Amino Acid Sequence (domain in AA coordinates: 62-106)
MVSVNPRPKGFPVFDSSNMSLPSSDGFSGIPATGRTSTVSFSEDPPTKIRKPYTIKKSRE
NWTDOEHDKFLEALHLFDRDWKKIEAFVSGKTVVQIRSHAQKYFLKVQKSGANEHLPLPR
PKRKASHPYPIKAPKNVAYTSLPSSSTLPLEPGYLYSSDSKSLMGNQAVCASTSSSWNH
ESTNLPKPVIEEPGVSATAPLPNNRCQEDTERVRAVTKPNNNEESCEKPHRVMPNFAEV
YSFIGSVFDPNTSGHLQRLKQMDPINMETVLLLMQNL SVNLTSPFEAEQRRLISSYSAKA
LK*

>G26 (73..729)
TTGGCTTGTAACCAACCCATCTTTGACTTCAAAAATAAAAATAAATCATAATTGA
CATCATCGGATAATGCATAGCGGGAAGAGACCTCTATCACCAGAATCAATGGCCGGAAAT
AGAGAAGAGAAAAAGAGTTGTGTTGTTGCTCAACTTTGTGCGGAATCTGATGTGTCTGAT
TTTGTCTCTGAACCTCACTGGTCAACCCATCCCATCATCCATTGATGATCAATCTTCGTCTG
CTTACTCTTCAAGAAAAAGTAACCTCGAGGCAACGAACTACAGAGGCGTGAGGCAAGA
CCGTGGGGAAAAATGGGCGGCTGAGATTCTGTGACCCGAACAAGGCAGCTCGTGTGTGGCTT
GGGACGTTTCGACACTGCAGAAGAAGCCGCCTTAGCGTATGATAAAGCTGCATTTGAGTTT
AGAGGTCACAAGGCCAAGCTTAACCTCCCGAGCATATTCTGTGTCAACCCTACTCAACTC
TATCCATCGCCCGCTACTTCCCATGATCGCAATTATCGTGACACCACCTAGTCCACCTCCA
CCAATTGCTCCTGACATACTTCTTGATCAATATGGCCACTTTCAATCTCGAAGTAGTGAT
TCCAGTGCCAACCTTGTCCATGAATATGCTGTCTTCTTCGTCTTCATCTTTGAATCATCAA
GGGCTAAGACCAAAATTTGGAGGATGGTGAAAACGTGAAGAACATTAGTATCCACAAACGA
CGAAAATAACATGTTAATGGCATAAATATCTCTTCTCGTCCAAGTTATCAAACGCATTGACC
TCCGGCTTTGATCATTTTAGGCGCTTAATCTCTTTACGACTTCATTTTGGTAGTCTTTAA
AGAGTCTATGGAGTGGATTTAGCTAGGAATCAGGCCTTATGGATGAAAAATATATAAATT
TTGAACATGACTATGCAAGAATGGGATGAAGACTACTTAGCTTGGAAAACGTCCTGATAG
GTCATGACGACTATATCCACAGAAGATGACCGACGGAGACAACAACATGCCTCACCTGAT
CGACCGATCAAATGAGATAATGTGTTGACCGGACCGGTCCGATCAGGTTGGGTTCGAGTAT
ATCA

>G26 Amino Acid Sequence (domain in AA coordinates: 67-134)
MHSGKRPLSPESMAGNREEKKELCCSTLSESDVSDVSELTGQPISSIDDQSSSLTLQ
EKSNSRQRNRYRGVRQRPWGKWAABEIRDPNKAARVWLGTFTABEAAALAYDKAAFEFRGHK
AKLNFPEHIRVNPTQLYPSPATSHDRIIVTPSPPPPIAPDILLDQYGHFQSRSSDSSAN
LSMNM LSSSSSLNHQGLRPNLEDGENVKNISIHKRRK*

>G263 (48..902)
TTTTTAGTTTTATTTTTCTGTGGTAAAAATAAAAAAGTTCCGCCGGAGATGACGGCTGTGA
CGGCGGCGCAAAGATCAGTTCCGGCGCCGTTTTTAAGCAAAACGTATCAGCTAGTTGATG
ATCATAGCACAGACGACGTGTTTCATGGAACGAAGAAGGAACAGCTTTTGTCTGTGGA
AAACAGCAGAGTTTGCTAAAGATCTTCTTCTCAATACTTCAAGCATAATAATTCTCAA
GCTTCATTTCGTCAGCTCAACACTTACGGATTTCGTAAAACTGTACCGGATAAATGGGAAT
TTGCAAAACGATTATTTCCGGAGAGGCGGGGAGGATCTGTTGACGGACATACGACGGCGTA
AATCGGTGATTGCTTCAACGGCGGGGAAATGTGTTGTTGTTGTTGTTGCTTCTGAGTCTA
ATTCTGGTGGTGGTGATGATCACGGTTCAAGCTCCACGTCATCACCCGGTTCTGTCGAAGA
ATCCTGGTTCCGTGGAGAACATGGTTGCTGATTATCAGGAGAGAACGAGAAGCTTAAAC
GTGAAAACAATAACTTGAGCTCGGAGCTCGCGGCGGCGAAGAAGCAGCGCGATGAGCTAG
TGACGTTCTTGACGGGTCATCTGAAAGTAAGACCGGAACAAATCGATAAAATGATCAAAG
GAGGGAAATTTAAACCGGTGGAGTCTGACGAAGAGAGTGAGTGCGAAGGTTGCGACGGCG
GCGGAGGAGCAGAGGAGGGGGTAGGTGAAGGATTGAAATTGTTGGGGTGTGGTTGAAAG
GAGAGAGAAAAAGAGGGGACCGGGATGAAAAAATTATGTGGTGAGTGGGTCCCGTATGA
CGGAAATAAAGAACGTGGACTTTCACGCGCGGTTGTGGAAAAGCAGCAAAGTCTGCAACT
AAAAAAGAGTAGAAGACTGTTCAAACCAGCGTGTGACACGTCATCGACGACGACGAAAA

AAATGATTTAAAAAACTATTTTTTCCGTAAGGAAGAAAAGTTATTTTTATGTTTTAAAA
AGGTGAAGAAGGTCCAGAAGGATCAACGCAATATATAAATGGATTTTCATGTATTATAT
AATTTAATTAGTGTATTAAGAAAA

>G263 Amino Acid Sequence (domain in AA coordinates: TBD)

MTAVTAAQRSVPAPFLSKTYQLVDDHSTDDVVSWNNEGTAFVWVKTAFAKDLLPQYFKH
NNFSSFIRQLNTYGFRTKVPDKWFEFANDYFRRGGEDLLTDIRRRKSVIASTAGKCVVVG
PSESNSGGDDHSGSSSTSSPGSSKNPGSVENMVADLSGENEKLKRENNLSSELAAAKKQ
RDELVTFLTGLHVKRPEQIDKMIKGGKFKPVESDEESECEGCDGGGGAEVGEGLKLF
VWLKGERKKRDRDEKNYVVSGRMTEIKNVDFHAPLWKSSKVCN*

>G308 (196..1794)

AGTAATTTAGTTTTTTTTTTTTTTTTTACAATTTATTTTGTATTAGAAGTGGTAGTGG
AGTGAAAAACAAATCCTAAGCAGTCCTAACCGATCCCCGAAGCTAAAGATTCTTCACCT
TCCCAAATAAAGCAAAACCTAGATCCGACATTGAAGGAAAAACCTTTTAGATCCATCTCT
GAAAAAAACCAACCATGAAGAGAGATCATCATCATCATCAAGATAAGAAGACTATG
ATGATGAATGAAGAAGACGACGGTAACGGCATGGATGAGCTTCTAGCTGTTCTTGTTAC
AAGGTTAGGTCATCGAAATGGCTGATGTTGCTCAGAACTCGAGCAGCTTGAAGTTATG
ATGTCTAATGTTCAAGAAGACGATCTTCTCAACTCGCTACTGAGACTGTTCACTATAAT
CCGGCGGAGCTTTACACGTGGCTTGATTCTATGCTCACCAGCTTAATCCTCCGTCGCTC
AACGCCGAGTACGATCTTAAAGCTATTCGCCGTGACCGGATTCTCAATCAGTTCGCTATC
GATTCGGCTTCTTCTGCTAACCAAGGCGGCGGAGGAGATACGTATACTACAAACAAGCGG
TTGAAATGCTCAAACGGCGTCTGGAACACCAACAGCGACGGCTGAGTCAACTCGGCAT
GTTGTCCTGGTTGACTCGCAGGAGAACGGTGTGCGTCTCGTTCACGCGCTTTTGCGTTGC
GCTGAAGCTGTTTCAAGAAGGAGAATCTGACTGTGGCGGAAGCTCTGGTGAAGCAATCGGA
TTCCTTAGCTGTTTCTCAATCGGAGCTATGAGACAAGTCGCTACTTACTTCGCCGAAGCT
CTCGCGCGGCGGATTTACCGTCTCTCTCCGTCGCAGAGTCCAATCGACCACTCTCTCTCC
GATACTCTTCAGATGCACCTTCTACGAGACTTGTCTTATCTCAAGTTCGCTCACTTCACG
GCGAATCAAGCGATTCTCGAAGCTTTTCAAGGGAAGAAAAGAGTTCATGTCAATTGATTTC
TCTATGAGTCAAGGCTCTCAATGGCCGGCGCTTATGCAGGCTCTTGCCTTCGACCTGGT
GGTCCCTCTGTTTTCCGGTTAACCAGGAATGGTCCACCGGCACCGGATAATTCGATTAT
CTTCATGAAGTTGGGTGTAAGCTGGCTCATTTAGCTGAGGCGATTACGTTGAGTTTGAG
TACAGAGGATTTGTGGCTAACACTTTAGCTGATCTTGATGCTTCGATGCTTGAGCTTAGA
CCAAGTGAGATTGAATCTGTTGCGGTTAACTCTGTTTTCGAGCTTCACAAGCTCTTGGGA
CGACCTGGTGCGATCGATAAGGTTCTTGGTGTGGTGAATCAGATTAAACCGGAGATTTTC
ACTGTGGTTGAGCAGGAATCGAACATAATAGTCCGATTTTCTTAGATCGGTTTACTGAG
TCGTTGCATTATTACTCGACGTTGTTTGACTCGTTGGAAGGTGTACCGAGTGGTCAAGAC
AAGGTCATGTCCGAGGTTTACTTGGGTAAACAGATCTGCAACGTTGTGGCTTGTGATGGA
CCTGACCGAGTTGAGCGTCATGAAACGTTGAGTCAGTGGAGGAACCGGTTCCGGTCTGCT
GGGTTTGCGGCTGCACATATGTTTCAATGCGTTTAAAGCAAGCGAGTATGCTTTTGCGT
CTGTTCAACGCGCGGTGAGGTTTATCGGGTGGAGGAGAGTGACGGCTGTCTCATGTTGGGT
TGGCACACACGACCGCTCATAGCCACCTCGGCTTGGAACTCTCCACCAATTAGATGGTG
GCTCAATGAATTGATCTGTTGAACCGGTTATGATGATAGATTTCCGACCGAAGCCAACT
AAATCCTACTGTTTTTCCCTTGTCACTTGTAAAGATCTTATCTTTCATTATATTAGGTA
ATTGAAAAATTTTAATCTCGCCTAAATTACT

>G308 Amino Acid Sequence (domain in AA coordinates: 270-274)

MKRDHHHHQDKKTMMNEEDDNGMDELLAVLGYKVRSEMADVAQKLEQLEVMSNVQ
EDDLSQLATETVHYNPAELYTWLDSMLTDLNPPSSNAEYDLKAIPGDAILNQFAIDSASS
SNQGGGDTYTTNKRLKCSNGVETTTATAESTRHVVLDVDSQENGVRVHALLACAEAVQ
KENLTVAEALVKQIGFLAVSQIGAMRQVATYFAEALARRIYRLSPSQSPIDHSLSDTLQM
HFTYETCPYLKFAHFTANQAILAEAFQGGKRVHVIDFSMSQGLQWPAQMQLALRPGPPVF
RLTGIGPPAPDNFDYDLHEVGCKLAHLAEAIHVEFEYRGFVANTLADLDASMLELRPSEIE
SVAVNSVFEHLHKLGRPGAIDKVLGVVNQIKPEIFTVVEQESNHNSPIFLDRFTESLHY
STLFDSLEGVPSGQDKVMSEVYLGKQICNVVACDGPDRVERHETLSQWRNRFSGAGFAAA
HIGSNAPKQASMLLALFNGGEGYRVEESDGLMLGWHTRPLIATSAWKLSTN*

>G38 (149..1156)

GAGGAAAACTCGAAAAAGCTACACACAAGAAGAAGAAAAGATACGAGCAAGAAGACT
AAACACGAAAGCGATTTATCAACTCGAAGGAAGAGACTTTGATTTTCAAATTCGTCCCC
TATAGATTGTGTTGTTTCTGGGAAGGAGATGGCAGTTTATGATCAGAGTGGAGATAGAAA

CAGAACACAAATTGATACATCGAGGAAAAGGAAATCTAGAAGTAGAGGTGACGGTACTAC
TGTGGCTGAGAGATTAAAGAGATGGAAAGAGTATAACGAGACCGTAGAAGAAGTTTCTAC
CAAGAAGAGGAAAGTACCTGCGAAAGGGTCGAAGAAGGGTTGTATGAAAGGTAAAGGAGG
ACCAGAGAATAGCCGATGTAGTTTCAGAGGAGTTAGGCAAAGGATTGGGGTAAATGGGT
TGCTGAGATCAGAGAGCCTAATCGAGGTAGCAGGCTTTGGCTTGGTACTTTCCCTACTGC
TCAAGAAGCTGCTTCTGCTTATGATGAGGCTGCTAAAGCTATGTATGGTCCTTTGGCTCG
TCTTAATTTCCCTCGGTCTGATGCGTCTGAGGTTACGAGTACCTCAAGTCAGTCTGAGGT
GTGTACTGTTGAGACTCCTGGTTGTGTTTCATGTGAAAACAGAGGATCCAGATTGTGAATC
TAAACCCCTTCTCCGGTGGAGTGGAGCCGATGTATTGTCTGGAGAATGGTGCAGGAGAGAT
GAAGAGAGGTGTTAAAGCGGATAAGCATTGGCTGAGCGAGTTTGAACATAACTATTGGAG
TGATATTCTGAAAGAGAAAGAGAAAACAGAAGGAGCAAGGGATTGTAGAAACCTGTGAGCA
ACAACAGCAGGATTTCGCTATCTGTTGCAGACTATGGTTGGCCCAATGATGTGGATCAGAG
TCACCTGGATTCTTCAGACATGTTTGTATGTCGATGAGCTTCTACGTGACCTAAATGGCGA
CGATGTGTTTGCAGGCTTAAATCAGGACCGGTACCCGGGGAACAGTGTGCGCAACGGTTC
ATACAGGCCCCGAGAGTCAACAAAGTGGTTTTGATCCGCTACAAAGCCTCAACTACGGAAT
ACCTCCGTTTTAGCTCGAGGGAAGGATGGTAATGGATTCTTCGACGACTTGAGTTACTT
GGATCTGGAGAACTAAACAAAACATATGAAGCTTTTTGGATTGTATTTGCCTTAATC
CCACAACGACTGTTGATTCTCTATCCGAGTTTTAGTGATATAGAGAACTACAGAACACGT
TTTTTCTTGTATAAAGGTGAAGTGTATATATCGAAACAGTGATATGACAAATAGAGAAGA
CAACTATAGTTTGTAGTCTGCTTCTCTTAAGTTGTTCTTTAGATATGTTTATGTTTTG
TAACAACAGGAATGAATAATACACACTTGTGAAGCTTTTAAAAAAAAAAAAAAAAAAAAA
>G38 Amino Acid Sequence (domain in AA coordinates: 76-143)
MAVYDQSGDRNRTQIDTSRKRKRSRSGDGTVAERLKRWKEYNETVEEVSTKKRKVPAKG
SKKGCМКGKGPPENSRCFSRQVRQRIWGKWVAEIREPNRGSRLWLGTFTAQEAASAYDE
AAKAMYGPLARLNFPRSDASEVTSTSSQSEVCTVETPGCVHVKTEDPDCESKPFSGGVEP
MYCLENAGAEEMKRGVKADKHWLSEFEHNYWSDILKEKEKQKEQGIVETCQQQQQDLSLVA
DYGWPNDVDQSHLDSSDMFVDLELLRLDNGDDVFAGLNQDRYPGNSVANGSYRPESQSG
FDPLQSLNYGIPPFQLEGKDGNGFFDDLSYLDLEN*

>G43 (38..643)
CTCCTGTCCTTGTCTAAAGAAAAAGAGAGAGGAAGAAATGGAGACTTTTGAGGAAAGCTC
TGATTTGGATGTTATACAGAAACATCTATTTGAAGACTTGATGATCCCTGATGGTTTCAT
TGAAGATTTTGTCTTTGATGATACTGCTTTTGTCTCCGACTCTGGTCTCTAGAACCCTT
TAACCCAGTTCCGAAACTGGAACCTAGTTCACCTGTTCTTGATCCAGATTCTATGTCCA
AGAGATTCTGCAAAATGGAAGCAGAAATCATCATCATCAACAACAACGTCACCTGA
GGTTGAGACTGTCTCAAACCGGAAAAAAACAAAGAGGTTTGAAGAAACGAGACATTACAG
AGGCGTGAGAAGGAGGCCATGGGGGAAATTTGCAGCAGAGATTGAGATCCGGCAAAGAA
AGGATCCAGGATTGAGTTAGGCCTTTTGAGAGTGATATGATGCTGCAAGGGCTTACGA
CTATGCAGCTTTTAAGCTCAGGGGAAGAAAAGCTGTTCTCAACTTTCTTTGGATGCCGG
AAAGTATGATGCTCCGGTCAATTTCATGCCGAAAAAGGAGGAGAACCGATGTACCACAGCC
TCAAGGAACAACAACAAGTACTTCATCATCGTCATCAACTAATGGGGGAATAGTGATGT
TTAATTAGTATATATAGGTTAATATCTTAAGTATGTGAAGCATCATGTATAGAGCCAAGA
ACCTGTTAGACTAGTGTAAGTAAAGAACTCTTGCAAAATATGTACTAAAGAGTTCCTGT
AACAATGGAACCTTCTGCGTTTTCTCTTGTCTTAAAGAGCTTAAGGTTCTAGAAACAAAGT
TCTTGTCTTTCCGGTTTAAA
AAAAAAAAAA

>G43 Amino Acid Sequence (domain in AA coordinates: 104-172)
METFEESDLDVIQKHLFEDLMIPDGFIEDFVDDTAFVSGLSLEFPNVPKLEPSSPV
LDPDSYVQEILOMEAESSSSSSTTSPEVETVSNRKKTKRFEETRHYRGVRRRPWGKFAA
EIRDPAKKGSRIWLGFESDIDAARAYDYAAFKLGRKAVLNFLDAGKYDAPVNSCRKR
RRTDVPQPQGTSTSSSSSN*

>G536 (1..768)
ATGTCGACAAGGAAGAGAATGTTTACATGGCGAAATTAGCCGAACAAGCTGAACGTTAC
GAAGAAATGGTTGAATTCATGGAGAAAGTTGCGAAAACCTGTTGATGTTGAGGAACTTTCA
GTTGAAGAGAGGAATCTTCTCTCTGTTGCTTACAAGAACGTGATTGGAGCGAGAAGAGCT
TCGTGGAGAATCATTTCTTCGATTGAGCAGAAAGAAGAGAGCAAAGGGAACGAAGATCAT
GTTGCTATTATCAAGGATTACAGAGGAGAGATTGAATCCGAGCTTAGCAAAATCTGTGAT
GGGATTTTGAATGTTCTTGAAGCTCATCTTATTCCTTCTGCTTACCAGCTGAATCTAAA

GTGTTTTATCTTAAGATGAAGGGTGATTATCATAGGTATCTTGCTGAGTTTAAGGCTGGT
GCTGAAAGGAAAGAAGCTGCTGAAAGCACTTTGGTTGCTTACAAGTCTGCTTCCGACATT
GCCACTGCTGAGTTAGCTCCTACTCACCGATAAGGCTTGGTCTTGCACTCAACTTCTCT
GTGTTTTACTATGAAATCCTCAACTCGCTGATCGTGCTTGACAGCTCGCAAAGCAGGCG
TTTGATGATGCAATCGCTGAGTTAGATACATTGGGGTGAGGAATCATACAAGGACAGTACA
CTGATTATGCAGCTTCTTAGAGACAATCTCACTCTCTGGACTTCAGATATGACTGACGAA
GCAGGAGATGAGATTAAGGAGGCATCAAAGCCCGATGGTGCCGAGTAA

>G536 Amino Acid Sequence (domain in AA coordinates:226-233)
MSTREENVYMAKLAEQAEYEEVMEFMEKVAKTVDVVEELSVVEERNLLSVAYKNVIGARRA
SWRIISSIEQKEESKGNEDHVAIIKDYRGEIESELSKICDILNVLEAHLIPSASPESK
VFYLMKMGDYHRYLAEFKGAERKEAAESTLVAYKSASDIATAELAPTHPIRLGLALNFS
VFYYEILNSPDRACSLAKQAFDDAIAELDTLGEESYKDSLIMQLLRDNLTLWTSDMTDE
AGDEIKKASKPDGAE*

>G567 (38..1273)

AAAAAGAAGAATCAGAAAGTGAAAAAGAGAGCGAGCGATGAACAGTATCTTCTCCATTGA
CGATTTCTCCGATCCTTTCTGGGAAACTCCTCCGATTCCTCTCAATCCCGACTCTTCTAA
GCTTGTTACGGCGGATGAAGTTAGCCAGAGTCAACCGGAATGGACTTTCGAGATGTTTCT
CGAAGAGATTTCTTCGTCGGCGGTGAGCTCTGAGCCACTTGGTAACAACAACACGCGAT
CGTCGGTGTTTCTTCGGCGCAATCTCTTCTTCTGTTTCCGGACAGAATGATTTTCGAGGA
TGATAGTCGATTTCTGATCGCGATTTCGGGAAATTTGGATTGTGCTGCTCCCATGACGAC
GAAGACGGTGAATGTTTGGAATTCGATGATTATCGTCGTGTTCTTAAGAACAAGCTTGAGGC
TGAGTGCAGGACTGGTGTTTCTTCTCGGGTTGGGTCTGTGAAGCCTGAAGATTCGACTAG
TTCTCCAGAACTCAACTTCAACCAGTTCAATCCAGTCTCTTACTCAAGGAGAACTTGG
TGTTACTTCTTCTTACCAGCTGAGGTGAAAAAACTGGTGTATCAATGAAGCAGGTTAC
TAGTGGATCGTCGAGAGAAATATTCTGATGACGAGGACCTTGATGAAGAGAATGAAACCAC
CGGTTCCCTTGAAGCCAGAGGACGTTAAAAAATCTAGAAGGATGCTGTCAAATCGTGAGTC
AGCTAGGCGATCTAGAAGGAGAAAGCAGGAGCAAAACAAGTGACCTCGAAACACAGGTTAA
TGATCTAAAAGGTGAGCATTCATCACTTCTTAAACAACAGGACATGAATCACAAGTA
TGACGAGGCTGCTGTTTGGAATAGAATACTAAAGGCTGACATTGAGACATTAAGAGCTAA
GGTGAAAAATGGCGGAAGAAACCGTGAAGAGAGTAACAGGAATGAATCCGATGCTTCTCGG
AAGATCAAGTGGACATAACAACAACAAGAGATGCCAATAACTGGTAACAACAGGATGGA
TTCTTCTAGCATTATTCAGCTTATCAACCACACTCAAACCTAAACCATATGTCAAACCA
AAACATCGGGATCCCAACCATCTACCTCCAAGACTCGGAAACAATTTTCGCTGCTCCTCC
ATCCCAAACAGCTCTCCTTGCAGAGAATTAGAAATGGGCAAAATCACCATGTTACTCC
AAGCGCAACCCGTATGGCTGGAATACCGAACCTCAGAACGATTGAGCATGGCCGAAAAA
ATGCGTGGAGTATCAACAAGAAGCGGTTTCGCACTATATTAATGTCTATGCATCTGT
AATTTGTAAGTGTTATTAGTTACGAATCATGAGAAAAACATCTTGTGAAAAATACAGTCTC
ATGGCTTATATATATATATAAGCTCTGTCTTATAACATTACAAGATTCTTATTGAGAAT
CGTCTTTCTATTTATAGCTAATAAAAAAAAAAAAAAAAAA

>G567 Amino Acid Sequence (domain in AA coordinates 210-270)
MNSIFSIDDFSDPFWETPPIPLNPDSSKPVTADEVSSQSQPEWTFEMFLEEISSSAVSSEP
LGNMNAIVGVSSAQSLPSVSGQNDFFEDDSRFRDRDSGNLDCAAPMTTKTVNVDSDDYRR
VLKNKLEAECATGVSLRVGSVKPEDSTSSPETQLQPVQSSPLTQGELGVTSSSLPAEVKKT
GVSMKQVTSGSSREYSDDDLDEENETGSLKPEDVKKSRRLSNRESARRSRRRKQEQT
SDLETQVNDLKGEHSSLLKQLSNMNHKYDEAAVGNRILKADIEFLRAKVKMAEETVKRV
GMNPMMLGRSSGHNNNNRMPITGNNRMDSSSIIPAYQPHSNLHNHMSNQNIPIILPPRL
GNNFAAPPSQTSSPLQIRNGQNHVTPSANPYGWNTPEQNDSAWPKKCVD*

>G680 (338..2275)

CAGTTATCTTCTTCTTCTTCTTCTGTTTTTAAATTTATTTTATAGAGAATTTTTTTTG
TTTTGCTTCCGATTTGATTATTTCCGGGAACGATGACTTCTCCGGGAGTTCCCGGTGAG
ATGATAAGTCAGATTGCATACTTGTCTCCTCCATGGCTACTCTCAAGGGTTTGGCTGCG
GTGGATTCTGTTTGGTTTCTCTAGAACTTAAAGAGGTTATCACAACGGCTTTGCAATTTGA
AACTTTTCATGTTTGGGGAGATCAAAGATGGTTTCTTTTTTATACTTTACTTGTAGAGA
GGATTTGAAGCAGCGAATAGCTGCAACCGGTCCTGTTATGGATACTAATACATCTGGAGA
AGAATTATTAGCTAAGGCAAGAAAGCCATATACAATAACAAGCAGCGAGAGCGATGGAC
TGAGGATGAGCATGAGAGGTTTCTAGAAGCCTTGAGGCTTTATGGAAGAGCTTGGCAACG
AATTGAAGAACATATTGGGACAAAGACTGCTGTTTCAGATCAGAAGTCATGCACAAAAGTT

CTTCACAAAGTTGGAGAAAGAGGCTGAAGTTAAAGGCATCCCTGTTTGCCAAGCTTTGGA
CATAGAAATTCCGCCTCCTCGTCCTAAACGAAAAACCAATACTCCTTATCCTCGAAAACC
TGGGAACAACGGTACATCTTCCTCTCAAGTATCATCAGCAAAAGATGCAAAACTTGTTC
ATCGGCCTCTTCTTCACAGTTGAATCAGGCGTTCTTGGATTTGGAAAAATGCCGTTCTC
TGAGAAAACATCAACTGGAAAAGAAAATCAAGATGAGAATTGCTCGGGTGTTCCTACTGT
GAACAAGTATCCCTTACCAACGAAACAGGTAAGTGGCGACATTGAAACAAGTAAGACCTC
AACTGTGGACAACGCGGTTCAAGATGTTCCCAAGAAGAACAAGACAAAGATGGTAACGA
TGGTACTACTGTGCACAGCATGCAAACTACCCCTTGGCATTTCACGCAGATATTGTGAA
CGGGAATATAGCAAAATGCCCTCAAAATCATCCCTCAGGTATGGTATCTCAAGACTTCAT
GTTTCATCCTATGAGAGAAGAACTCACGGGCACGCAAATCTTCAAGCTACAACAGCATC
TGCTACTACTACAGCTTCTCATCAAGCGTTTCCAGCTTGTTCATTACAGGATGATTACCG
TTCGTTTCTCCAGATATCATCTACTTTCTCCAATCTTATTATGTCAACTCTCCTACAGAA
TCCTGCAGCTCATGCTGCAGCTACATTGCTGCTTGGTCTGGCCTTATGCGAGTGTGCG
GAATTCTGGTGATTCATCAACCCCAATGAGCTCTTCTCTCCAAGTATAACTGCCATTGC
CGCTGTACAGTAGCTGCTGCAACTGCTTGGTGGGCTTCTCATGGACTTCTTCTGTATG
CGCTCCAGCTCCAATAACATGTGTTCCATTCTCAACTGTTGCAGTTCCAACTCCAGCAAT
GACTGAAATGGATACCGTTGAAAATACTCAACCGTTTGAGAAAACAAACACAGCTCTGCA
AGATCAAACCTTGGCTTCGAAATCTCCAGCTTCATCATCTGATGATTGAGATGAGACTGG
AGTAACCAAGCTAAATGCGGACTCAAAACCAATGATGATAAAATTGAGGAGGTTGTGT
TACTGCCGCTGTGCATGACTCAAACTGCCCAGAGAAAATCTTGTGGACCGCTCATC
GTGTGGCTCAAATACACCTTCAGGGAGTGACGCAGAACTGATGCATTAGATAAAATGGA
GAAAGATAAAGAGGATGTGAAGGAGACAGATGAGAATCAGCCAGATGTTATTGAGTTAAA
TAACCGTAAGATTAAATGAGAGACAACAACAGCAACAACAATGCAACTACTGATTCGTG
GAAGGAAGTCTCCGAAGAGGGTCTATAGCGTTTCAGGCTCTCTTTGCAAGAGAAAGATT
GCCTCAAAGCTTTTCGCCTCCTCAAGTGGCAGAGAATGTGAATAGAAAACAAAGTGACAC
GTCAATGCCATTGGCTCCTAATTTCAAAGCCAGGATTCTTGTGCTGCAGACCAAGAAGG
AGTAGTAATGATCGGTGTTGGAACATGCAAGAGTCTTAAACGAGACAGACAGGATTAA
GCCATACAAGAGATGTTCAATGGAAGTGAAAGAGAGCCAAAGTTGGGAACATAAACAATCA
AAGTGATGAAAAAGTCTGCAAAAGGCTTCGATTGGAAGGAGAAGCTTCTACATGACAGAC
TTGGAGGTAAAAAAAACATCCACATTTTATCAATATCTTTAAATCTAGTGTAGTAG
TTTGCTTCTCCAATCTTTATGAAAGAGACTTTTAATTTCTTCCGAACATTTCTTTGGT
CATGTCAGGTTCTGTACCATATTACCCCATGTCTTGTCTCTTGTCTCTGTTGTATGC
TACTTGTGGTCTATATGTCATCTGCTACTACTGTTAATTAACCATTAGCAATGGATTG
TCTTTA

>G680 Amino Acid Sequence (domain in AA coordinates: 24-70)
MDTNTSGEELLAKARKPYTITKQERWTEDEHERFLEALRLYGRAWQRIEEHIGTKTAVQ
IRSHAQKFFTKLEKEAEVKGIPVCQALDIEIPPPRKRKPNTPYPRKPGNNGTSSSQVSS
AKDAKLVSASSSQLNQAFDLEKMPFSEKTSTGKENQDENCSTVSVNXYPLPTKQVSG
DIETSKTSTVDNAVQDVPKKNKDKDNDGTTVHSMQNYPWHFHADIVNGNIAKCPQNHPS
GMVSQDFMFHMPREBETHGHANLQATTASATTTASHQAFPACHSQDDYRSFLQISSTFSNL
IMSTLLQNPAHAHAATFAASVWPYASVGNSGDSSTPMSSSPPSITAIAAATVAAATAWWA
SHGLLPVCAPAPITCVPFSTVAVPTPAMTEMDTVENTQPFQKNTALQDQTLASKSPASS
SDDSDGTGVTCLNADSKTNDKIEVVVTAAVHDSNTAQKKNLVDNRSSCGSNTPSGSDAE
TDALDKMEKDVEDVKETDENQPDVIELNNRKKIMRDNNNNNATTDWKEVSEEGRIAFQ
ALFARERLPQSFSPPQVAENVNRKQSDTSMPLAPNFKSQDSCAADQEGVVMIGVGTCKSL
KTRQTGFKPKRCSMEVKESQVGNINNQSDEKVKRLRLEGEAST*

>G867 (64..1098)

CACAACACAAAACATTTCTGTTTTCTCCATTGTTTCAAACCATAAAAAAACACAGAT
TAAATGGAATCGAGTAGCGTTGATGAGAGTACTACAAGTACAGGTTCCATCTGTGAAACC
CCGGCGATAACTCCGGCGAAAAAGTCGTCGGTAGGTAACTTATACAGGATGGGAAGCGGA
TCAAGCGTTGTGTTAGATTACAGAGAACGGCGTAGAAGCTGAATCTAGGAAGCTTCCGTCG
TCAAAATACAAAGGTGTGGTCCCAACCAACGGAAGATGGGGAGCTCAGATTTACGAG
AAACACCAGCGGTGTGGCTCGGGACATTCAACGAAGAAGACGAAGCCGCTCGTGCCTAC
GACGTCGCGGTTACAGGTTCCGTCGCCGTGACGCCGTCAAAATTTCAAAGACGTGAAG
ATGGACGAAGACGAGGTCGATTCTTGAATTCTCATTCGAAATCTGAGATCGTTGATATG
TTGAGGAAACATACTTATAACGAAGAGTTAGAGCAGAGTAAACGGCGTCGTAATGGTAAC
GGAAACATGACTAGGACGTTGTTAACGTCGGGGTTGAGTAATGATGGTGTTCACGACG

GGGTTTAGATCGGCGGAGGCACTGTTTGAGAAAGCGGTAACGCCAAGCGACGTTGGGAAG
CTAAACCGTTTGGTTATACCGAAACATCACGCAGAGAAACATTTTCCGTTACCGTCAAGT
AACGTTTCCGTGAAAGGAGTGTGTTGAACTTTGAGGACGTTAACGGGAAAGTGTGGAGG
TTCCGTTACTCGTATTGGAACAGTAGTCAGAGTTATGTTTGTACTAAAGGTTGGAGCAGG
TTCGTTAAGGAGAAGAATCTACGTGCTGGTGACGTGGTTAGTTTCAGTAGATCTAACGGT
CAGGATCAACAGTTGTACATTGGGTGGAAGTCGAGATCCGGGTCAGATTTAGATGCCGGT
CGGGTTTTGAGATTGTTCCGAGTTAACATTTACCGGAGAGTTCAAGAAACGACGTCGTA
GGAAACAAAAGAGTGAACGATACTGAGATGTTATCGTTGGTGTGTAGCAAGAAGCAACGC
ATCTTTACGCCTCGTAACAACCTCTTCTTTTCTTTTCTTTTGTGTTTTAATAATTT
TTAAAACTCCATTTTCGTTTTCTTTATTTGCATCGGTTTCTTCTTCTTGTGTTTACCAA
GGTTCATGAGTTGTTTTGTTGTATTGATGAACTGTAAATTTTATTTATAGGATAAAATTT
TAAAAAAAAAAAAAAAAAAAAA

>G867 Amino Acid Sequence (domain in AA coordinates: 59-124)
MESSSVDESTTSTGSI CETPAITPAKSSVGNLYRMGSGSSVLDSENGVEAESRKLPS
KYKGVVPQPNRGRWGAQIYEKHQRVWLGT FNEEDEAARAYDVAVHRFRRRDAVTNFKDVKM
DEDEVDFLNSHSEIVDMLRKHTYNEELEQSKRRRNGNGNMTRLLTSGLSNDGVSTTG
FRSAEALFEKAVTPSDVGKLNRLVIPKHAEKHFPLPSSNVSVKGVLLNFEDVNGKVWRF
RYSYWNSSQSIVLTGWSRFVKEKNLRAGDVVSFSRSNGQDQQLYIGWKSRSRSGSDDLDA
GRVLRFLFVNISPESRRNDVVGNKRVNDTEMLSLVCSKKQRIFHAS*

>G956 (1..840)
ATGGAGGAGACAGAAAAGAATAAGGGCAGCATAAGTATGGTTGAGGCTAATCTACCTCCT
GGTTTATAGATTCCATCTAGAGACGACGAGCTCGTCTGTGACTACTTAATGAGAAGAACC
GTTCCGAGCCTCTATCAACCAGTTGTCTTGATCGACGTCGATCTTAACAAATGCGAGCCT
TGGGACATTCTCAACGGCGAGAGTGGGAGGGAAAGAATGGTACTTTTACAGCCAAAA
GACCGTAAATACGCAACAGGCTACAGAACAAACGGGCTACGCCACCGGTTATTGGAAA
GCCACCGGGAAAGATAGAGCAATCCAAGAAACGGTGGTCTTGTGGGTATGAGAAAGACA
CTTGTGTTTTACCGAGGTGCGATCCCCTAAAGGTCGTAAACTGATTGGGTGATGCATGAG
TTTCGTCTCCAAGGAAACTTCTTACCCTCCCTAATTCTCTCGAGGAAGAGTGGGTA
TTGTGTAGAGTTTTCCACAAGAACAGCAACGAGCTGATATAGACGACATCAACAAGGAGC
TGCTCTGATGCAACAGCTTCTGCAATTCATGGACTCTTACATCAACTTCGACCATCATCAC
ATCATCAATCAGCATGTACCCTGCTTCTCCAATAATTTGTACATAACCAAAACCAACCA
TCCGGTTTAACTCCAAGAACTCCAGCCCATTTGTTTAAATGCTTCCCCTGATCAAATGATT
CTCAGAACTTTGCTAAGTCAACTCACAACAAAAGTCAAGAATCACAGAGTCGTGGAGAC
GGAAGCTCAGAGAGCCAAATGACCGACATTGGCATCCCAAGCCATGCATGGAATTACTGA

>G956 Amino Acid Sequence (domain in AA coordinates: TBD)
MEETKKNKGSISMVEANLPPGFRFHPRDDELVC DYLMRRTVRSLYQPVLIDVDL NKCEP
WDIPQ TARVPKGEWFYFSQKDRKYATGYRTNRATGYWKATGKDRAIQ RNGGLVGM RKT
LVFYRGRSPKGRKTDVWMHEFRLQ GKLLHHS PNLSEEWVL CRV FHKNSNGADIDDI TRS
CSDATASAFMDSYINFDH HHI INQHVP CFSNNLSHNQTNQ SGLISKNSSPLFNASPDQMI
LRTL LSQLTKKVEESQSRGDGSS ESQ LTDIGIP SHAWNY*

>G996 (53..1063)
CGATCGATCTTGAATTGATCTTTGTAGTATTTTATTTACATATATATATAGATGGGAAG
ACATTCATGTTGTTACAAACAGAAACTGAGGAAAGGACTTTGGTCTCCTGAAGAAGATGA
GAAGCTTCTTCGTTACATCACTAAGTATGGTCATGGTTGCTGGAGCTCTGTCCCTAAACA
AGCTGGTTTACAGAGATGTGGAAGAGTTGTAGATTAAGATGGATAAATTTATTTAAGACC
AGATTTGAAGAGAGGAGCATTTTCTCAAGATGAAGAAAATCTCATTATTGAACTTCATGC
CGTTCTTGGCAATAGATGGTCTCAGATAGCTGCACAGCTTCTTGAAGAACCGACAATGA
AATCAAGAATCTTTGGAATCTTGTGTTGAAGAAGAAATTGAGGCTGAGAGGAATTGACCC
GGTTACACACAAGCTCTTAACCGAAATCGAAACCGGTACAGATGACAAAACAAACCGGT
TGAGAAGAGTCAACAGACCTACCTCGTTGAGACTGATGGCTCCTCTAGTACCACTACTTG
TAGTACTAACCAAAACAACAACACTGATCATCTTTATACCGGAAATTTCCGGTTTTCAACG
GTTAAGTCTAGAAAACGGTTCAAGAATCGCAGCCGGTCTGACCTCGGTATCTGGATTCC
CCAAACCGGAAGAACCATCATCATGTCGATGAAACCATCCCTAGTGCAGTGGTACT
ACCGGTTCAATGTTCTCATCCGGTTTAAACCGTTATAGATCCTCCAATCTCGGTTTAAAT
TGAATTGGAAGAACTCATTCTCAACCGGGCCAATGATGACAGAGCATCAGCAAATTCAGA
GAGTAACTACAACAATTCAACATCTTTGGAATGGGAATCTGAATTGGGGATTAACAAT
GGAGGAAAATCAAATCCATTCACAATATCGAATCATTCAAATTCGTCTTATACAGTGA

TATAAAATCAGAGACCAATTTTGGCACAGAGGCTACAAATGTTGGTATGTGGCCATG
TAACCAGCTTCAGCCTCAGCAACATGCATATGGCCATATATAAATCTTCTGTATATTAT
AA

>G996 Amino Acid Sequence (domain in AA coordinates: 14-114)
MGRHSCCYKQKLKGLWSPEEDEKLLRYITKYGHGCWSSVPKQAGLQRCGKSCRLRWINY
LRPDLKRGAFSQDEENLIIELHAVLGNRWSQIAAQLPGRTDNEIKNLWNSCLKKLRLRG
IDPVTHKLLTEIETGTDDKTKPVEKSQQTLYLVETDGSSTTTCTSTNNNDHLYTGNFG
FQRLSLENGSRIAAGSDLGWIWPQTGRNHHHHVDETIPIPAVVLPGSMFSSGLTGYRSSNL
GLIELENSFSTGPMTEHQIQESNYNNSTFFGNGNLNWGLTMEENQNPFITSNHSNSSL
YSDIKSETNFGTEATNVGMWPCNLQPPQHAYGHI*

>G1946 (90..1547)

TCTCACCTATTGTAAAAATCACCAGTTTCGTATATAAAACCCTAATTTTCTCAAAATTC
CAAAATATTGACTTGAATCAAAAATCCGAATGGATGTGAGCAAAGTAACCACAAGCGACG
GCGGAGGAGATTCAATGGAGACTAAGCCATCTCCTCAACCTCAGCCTGCGGCGATTCTAA
GTTCAAACGCGCCTCCTCCGTTTCTGAGCAAGACCTATGATATGGTTGATGATCACAATA
CAGATTGATGTTCTCTTGGAGTGCTAATAACAACAGTTTATCGTTTGGAAACACCGG
AGTTCTGCTCGCGATCTTCTTCTAAGAACTTTAAGCATAATAATTTCTCCAGCTTCGTTA
GACAGCTTAATACCTATGGTTTCAGGAAGGTTGACCCAGATAGATGGGAATTTGCGAATG
AAGGTTTTTTAAGAGGTGAGAAGCATTGCTACAATCAATAACTAGGCGAAAACCTGCCC
ATGGACAGGGACAGGGACATCAGCGATCTCAGCACTCGAATGGACAGAACTCATCTGTTA
GCGCATGTGTTGAAGTTGGCAAATTTGGTCTCGAAGAAGAAGTTGAAAGGCTTAAAAGAG
ATAAGAACGTCCTTATGCAAGAAGCTCGTCAGATTAAGACAGCAGCAACAGTCCACTGATA
ACCAACTTCAAACGATGGTTTCCAGCGTCTCCAGGGCATGGAGAATCGGCAACAACAATTAA
TGTCATTCCTTGCAAAGGCAGTACAAAGCCCTCATTTTCTATCTCAATTCTTACAGCAGC
AGAATCAGCAAAACGAGAGTAATAGGCGCATCAGTGATACCAGTAAGAAGCGGAGATTCA
AGCGAGACGGCATTGTCCGTAATAATGATTCTGCTACTCCTGATGGAAGATAGTGAAGT
ATCAACCTCCAATGCAGCAGCAAGCCAAAGCAATGTTTAAACAGCTTATGAAGATGGAAC
CTTACAAAACCGCGATGATGGTTTCTTCTAGGTAATGGTACGTCTACTACCGAGGGAA
CAGAGATGGAGACTTCATCAAACCAAGTATCGGGTATAACTCTTAAGGAAATGCCTACAG
CTTCTGAGATACAGTCATCACCATTGAAACAACTCCTGAAAATGTTTCCGGCAGCAT
CAGAAGCAACCGAGAACTGTATTCTTCCCTGATGATCTAACTCTTCCCGACTTCACTC
ATATGCTACCGGAAAATAATTAGAGAAGCCTCCAGAGAGTTTCATGGAACCAACCTGG
GAGGTTCTAGTCCATTACTAGATCCAGATCTGTTGATCGATGATCTTTGTCCTTCGACA
TTGACGACTTTCCAATGGATTCTGATATAGACCCTGTTGATTACGGTTTACTCGAACGCT
TACTCATGTCAAGCCCGGTTCCAGATAATATGGATTCAACACCAGTGGACAATGAAACAG
AGCAGGAACAAAATGGATGGGACAAAACTAAGCATATGGATAATCTGACTCAACAGATGG
GTCTCTCTCTCTGAAACCTTAGATCTCTCAAGGCAAAATCCTTGATTTTGGGAGTTT
TAAAGTCTTTTGAAGTAACACAGTCCCTGAGAGCAGCATATTCAT

>G1946 Amino Acid Sequence (domain in AA coordinates: 32-130)
MDVSKVTTSDGGGDSMETKPSPPQPAAILSSNAPPPFLSKTYDMVDDHNTDSIVWSAN
NNSFIWVKPPEFARDLLPKNFKHNNFSSFVRQLNTYGFVRKVPDRWEFANEGFLRGQKHL
LQSIITRRKPAHQGGQGHQHSQHSNGQNSSVSACVEVGKFGLEEEVERLKRDKNVLMQELV
RLRQQQSTDNQLQTMVQRLQGMENRQQQLMSFLAKAVQSPHFLSQFLQQNQNESNRR
ISDTSKKRRFRKRDGIVRNND SATPDGQIVKYQPPMHEQAKAMFKQLMKMEPYKTGDDGFL
LGNGTSTTEGTEMETSSNQVSGITLKEMPTASEIQSSSPIETTENVSAASEATENCIPS
PDDLTLPDFTHMLPENNSEKPPESFMEPNLGGSSPLLDPDLLIDDSLSFDIDDFPMDSDI
DPVDYGLLERLLMSSPVPDNMDSTPVDNETEQEQNGWDKTKHMDNLTQQMGLLSPETLDL
SRQNP*

>G217 (84..2618)

cttcgttcttaccgagttccacgagcattagcttcagagaccttgaaattggagtgcggtt
ggatcaaaaacagttgagcgaagatgaggattatgattaagggaggtgtttggaagaaca
ccgaagatgagattctcaagccgcctgatgaagtatggaagaaccaatgggctcgga
tctcgtctcttctcgttcgtaagtctgctaacaagtgtaaagctcgctggtagagtggtc
tcgatccatctatcaaaaagactgaatggaccagagaagaagatgagaagcttctacatc
ttgctaactctctgcctactcaatggagaactattgctcctattgtgggtcgtagacccat
ctcaatgtcttgagaggtatgagaagctccttgatgcagcatgcactaaggatgaaaatt
atgatgcagcgatgatccacgaaaattacgtcctggtagattgatccgaaccagaag

caaagcctgctcgtcctgatccggtagacatggacgaagatgagaaagaaatgctttctg
aagcaagagctagattggctaacacgaggggaaagaaggctaaaagaaaagctagagaaa
aacaacttgaggaagctagaaggcttctctctgcaaaaaagaagagaactaaaagcag
ctgggattgatggaaggcataggaagaaagagaagggaatcgactataatgcagaaa
ttccttttgaagaggcgacctgcgggattttatgatactgcggatgaagatcgtcctg
ctgatcaagtaaaatttccaactaccattgaagaacttgaaggaaaaagaagagctgatg
tagaagcacatttacgcaaaacagatgttgcaaggaataaaattgctcagagacaggatg
ctccagcagctatatgtgaagcaaacagctgaatgatccggaagttgtaggaagaggt
caaagctgatgttaccaccaccgcagatttcagaccacgagctagaagaaattgctaaga
tgggctatgccagtgccttcttgccgagaatgaggagctaacagaaggcagtgctgcta
ctcgtgcacttttgcaaaattactcacaacaccaaggcaaggatgacacccatgagga
cacctcaaagaactcctgctggttaaaggatgctattatgatggaagcagaaaacctgg
ccagattaagagactctcagacacctttgctaggaggagaaaatcctgagttgcaccctt
ctgacttcactggggctcactccgagaaagaaggagattcaaacgcctaataccaatgttga
cccttcaatgactcctggtggtgctggtcttactccaagaattggcttgacgccatcaa
gggatgggtcttcttttctatgacacccaaaggactcccttcagggatgaactcaca
ttaacgaagacatggacatgcagcaaatgcaaaacttgagaggcagagacgagaggaag
ctagaaggagtttacgctctggtttgactgggcttcctcagccaaagaacgagtacaaa
tagttgcacaacctcctcctgaggaaagtgaagagccagaagagaaaattgaggaagaca
tgtcagacaggatagcaggggaaaggcgaggagaagaagcaagacaacaggcattgctta
agaagagatccaaggtcttcagagagatcttctagaccccgagctgcttcattggcag
taattaggaactcgttgcttccagctgatggagacaaaagttctgtgttctcctactc
cgattgaggttgagataaaaatggtaagagaggagcttctacagttgctggagcatgata
atgcaaagtatccgcttgatgacaaagctgagaagaagaaggagccaagaaccgtacca
accgttctgcttctcaagttcttgcaattgacgattttgatgaaaatgagctccaagagg
ctgacaaaatgataaaggaggagggggaagttctgtgtgtgtcaatgggacatgagaaca
agacacttgatgattttgtagaagctcacaacacatgcgtgaatgatctcatgtatttcc
ccactcgaagcgcttacgagctctcaagtggtgctgggaacgcggacaaagttgcagctt
ttcaggaggagatggagaatgtgagaaaaagatggaggaggatgagaagaaggcagaac
acatgaaggccaagtacaaaacttatcaaaagggtcatgagaggaggcagagaccgtgt
ggacccaaatagaggcgacattgaagcaggctgagattgggtggaacagaagtagagtgt
ttaaagcattgaagaggcaagaagagatggctgcatcttttaggaaaaagaatttgcaag
aggaagtgataaagcaaaaggaaacagagagtaaatgcagactcgctatgggaatatgt
tggcaatgggtgaaaaagcagaggagataatgggtcggtttccgagcacaggcattgaaga
aacaagaggatgttgaagattctcacaactgaaagaagctaagctagccactggagagg
aagaggacatagccatagccatggaagcttctgcataaaaacttgagtttgtattgctt
acaagttttaaggagacgtagcttgactttgtattggtaagtttttttaatatgagtc
gactttgtaaaaagggttatgatattctctgttgtatgctttgcaagagtcaagaat
ttgaatgcttcaggatcaaaaaaaaaaaaaaaaaaaaaaaaaaaaaa

>G217 Amino Acid Sequence (conserved domain in AA coordinates: 8-67)

MRIMIKGGVWKNTEDEILKAAMKYGKNQWARISSLLVRKSAKQCKARWYEWLDPSIKKT
EWTREDEKLLHLAKLLPTQWRTIPIVGRTPSQCLERYEKLDDAACTKDENVDAADDP
KLRPGEIDPNPEAKPARPDPVMDDEDEKEMLSEARARLANTRGKKAKRKAREKQLEEARR
LASLQKRRELKAAGIDGRHRKRKRKIDYNAEIPFEKRAPAGFYDTADEDRPADQVKFPT
TIEELEGRKRRADVEAHLRKQDVARNKIAQRQDAPAILQANKLNDPEVVRKRSKMLMLPPP
QISDHELEELIAKMGYASDLLAENEELTEGSAATRALLANYSQTPRQGMTPMRTPQRTAG
KGDAIMMEAEENLARLRDSQTPLLGGENPELHPSDFGTGVTPRKKEIQTPNPMLTPSMTGG
AGLTPRIGLTPSRDSSFSMTPKGTFPRDELHINEDMDMQQSALKERQRRBEARRSLRSG
LTGLPQPKNEYQIVAQPPPEESEPEEKIEEDMSDRIAREKAEERQQAALLKKRSKVLQ
RDLPRPPAASLAVIRNSLLSADGDKSSVPPPTPIEVADKMVREELLQLEHDNAKYPLDD
KAEMKKGAKNRTNRSASQVLAIDDFDENELQEADKMIKEEGKFLCVSMGHENKTLDDFVE
AHNTCVNDLMYFPTRSAYELSSVAGNADKVAAFQEEMENVRRKKMEDEKKAHEMKAKYKT
YTKGHERRAETVWTQIEATLKQAEIGGTEVECFKALKRQEEMAASFRKKNLQEEVIKQKE
TESKLQTRYGNMLAMVEKAEIIMVGFRAQALKQEDVEDSHKLKEAKLATGEEEDIAIAM
EASA*

>G2192 (92..2971)

CGGAAAGAGATCAACCAACGATAGAGGAGAAGAAGAACTTGCATACGCAAAAAAAGTTTC

CCGGGAAATTCAGAACTGCTTTGGAAAAATGTGCGAGCCCGATGATAAATCCGCTAG
AAACGGCGTCACTACTCAACCTTCGAGGTCAAGGGAGCTTCTAATGGATGTTGACGACTT
AGATCTTGACGGTTCATGGCCACTAGATCAAATCCCTTACTTATCCTCATCGAATCGCAT
GATTTCTCCGATTTTTGTCTCCTCTTCTCTGAGCAGCCTTGCTCGCCTCTCTGGGCTTT
CTCCGACGGTGGAGGAAATGGTTTTCCACCACGCAACCTCCGGTGGCGATGATGAGAAGAT
CAGCTCTGTCTCCGGTGTTCCTTCTTTCGTCTCGCCGAGTATCCTCTCTTCTCCCTTA
CTCTTCTCCATCAGCAGCTGAGAACACAACAGAGAAGCATAACAGTTTCCAGTTTCCGTC
TCCATTGATGAGCCTAGTCCCACCAGAGAACAACAGACAACCTACTGTGTGATCAAAGAGAG
GATGACTCAGGCGCTTCGATACTTCAAAGAATCAACCGAACAAACAGTTTGGCTCAGGT
CTGGGCTCCTGTGAGAAAGAATGGTCGTGATTGCTGACGACTTTGGGTCAACCTTTTGT
TCTTAATCCTAATGGTAATGGGCTTAATCAATACAGGATGATCTCTCTCACATATATGTT
TCTGTGGATAGTGAAAGTGACGTAGAGCTCGGACTCCCGGGTCGAGTTTTTCCGTGAGAA
ATTGCCTGAATGGACTCCAAATGTTTCACTACTATTCCAGCAAAGAATCTCGCGGCTTGA
TCACGCCCTTGCACTACAACGTGCGTGGTACACTGGCCTTGCTGTCTTTAATCCCTCTGG
TCAGTCTGTCATAGGTGTTGTGGAACCTATAATGACCTCAGAGAAGATTCACTATGCACC
CGAAGTGGACAAAGTTTGCAAAGCCCTTGAGGCGGTAAATCTGAAAAGCTCGGAAATACT
TGATACCAAACAACACAGATATGCAATGAGAGTCGCCAAAACGCGCTTGCTGAGATTCT
CGAAGTGTGACAGTTGTATGTGAGACCCATAACTTGCCTCTCGCTCAGACTTGGGTTC
ATGTGACGATGGGAGCGTTCTTGCCAATGGTGGCGGTCTAAAGAAAACTGCACCAGCTT
TGACGGTAGCTGCATGGGTCAAATCTGCATGTCTACAACCGACATGGCCTGCTATGTCTG
GGATGCTCATGTCTGGGGCTTTAGAGATGCCTGTCTTGAACACCATCTCCAGAAAGGCCA
GGGAGTCGCTGGACGAGCTTTTCTCAATGGTGGCTCATGTTTCTGCAGAGACATCAACAA
GTTCTGCAAAACGCAGTACCCACTAGTCCATTATGCGCTCATGTTCAAGTTGACCACTTG
TTTTGCAATATCTCTCCAGAGCTCTTACACGGGCGACGACAGTTACATTCTTGAATTTTT
TCTTCTTCCGAGTATAACAGACGACCAAGAGCAAGATTGCTGTTGGGTTCATTTTTGGT
GACAATGAAAGAACATTTTCAGAGTCTGAGGGTTCATCTGGGGTTGACTTTGGTGAAGA
TGACGACAAATGTCTTTTCAGAGATCATCAAGCATTACCGGACAAGAAGGTTCAATCAAA
AATAGAATCCATTCGAGTTCCCTTTTCTGGTTTTAAGTCAAATGCAACAGAGACGATGTT
GATTCCTCAGCCTGTGGTTTCACTCTTCTGATCCAGTAAATGAGAAAAATCAACGTGGCCAC
TGTTAACGGTGTGGTTAAGGAGAAGAAGAAAAACAGAGAAAAAGCGTGGGAAGACTGAGAA
AACAAATCAGTCTAGATGTACTTCAGCAGTATTTCACTGGAAGTCTCAAAGACGCTGCAAA
GAGCCTAGGAGTTTGCCCCGACGACAATGAAGCGAATTTGACGGCAACACGGAATCTCGCG
GTGGCCATCGAGGAAGATCAAGAAAGTGAATCGTTCAATCACAAGCTGAAACGAGTCAT
CGAATCTGTTCAAGGTACTGATGGAGGCCCTCGACCTGACTTCCATGGCCGTTAGTTCCAT
CCCTTGGACACACGGTCAAACATCAGCACAGCCACTAACTCACCCTAATGGTTCCAAACC
ACCTGAGCTACCAACACCAATAATTCACTAACCTAAGTGGTCAAGTGATCACAGTCCGAA
CGAGCCAAATGGTTTCGCTGAGTTACCAACCAAGCAATGGTCAACAGCGATCACGAACGGT
GGATGAGAGCGCTGGGACTCCAACCTCTCATGGCTCATGTGACGGTAACCAATTAGATGA
ACCGAAAGTCCCAAATCAAGATCCGCTCTTACGGTTGGTGGATCACCCGGGCTCCTTTT
TCCACCTTATTCTAGAGATCATGATGTATCTGCAGCTTCTTCGCAATGCCGAACAGGCT
TCTTGGTTCTATAGACCATTTCGAGGAATGCTCATTGAAGACGCTGGAAGTTCAAAAGA
TCTGAGAAATCTCTGCCCCACTGCAGCATTGACGATAAGTTTCAAGACACAACTGGAT
GAACAATGATAATAATAGCAACAACAATTTATACGCTCCCCAAAGGAAGAGGCCATTGC
AAATGTTGCATGCGAACCATCAGGCTCAGAAATGAGAACGGTAACAATCAAAGCAAGTTA
CAAAGACGACATAATACGGTTTCAAGATATCCTCGGGTTTCAAGTATAATGGAATTGAAGGA
TGAAGTGGCTAAGAGGCTGAAAGTTGATGCAGGAACGTTTCGATATCAAGTATCTTGACGA
TGATAACGAATGGGTTTTAATAGCTTGTGATGCTGATCTTCAAGAATGTCTCGAGATCCC
TAGATCCTCCCGCAGGAAATCGTAAGGCTCTTAGTTTCATGATGTAACGACAAATCTAGG
GAGCTCCTGCGAGAGCACTGGAGAAATTGTGACCTGATAATTCATTGCAACTCTTTTGTA
ATAG

>G2192 Amino Acid Sequence (conserved domain in AA coordinates:600-700)

MCEPDDNSARNVGTTPPSRSRELLMDVDDLDDGWSPLDQIPYLSSSNRMISPIFVSSSS
EQPCSPWLWAFSDGGNGFHHATSGDDEKISSVSGVPSFRLAEYPLFLPYSSPSAAENTT
EKHNSFQFPSPLMSLVPPENTDNYCVIKERMTQALRYFKESTEQHVLAQVWAPVRKNGRD
LLTTLGQPFVLNPNNGNLNQYRMISLTYMFSVDSSESDVELGLPGRVFRQKLPEWTPNVQY
YSSKEFSRLDHALHYNVRGTLALPVFNPSGQSCIGVVELIMTSEKIHYAPEVDKVCKALE
AVNLKSSEILDHQTQICNESRQNALAEILEVLTVCETHNLPLAQTWVPCQHGSVLANG

GGLKKNCTSFDDGSCMGQICMSTTDMACYVVDAAHVWGFRDACLEHHLQKGQGVAGRAFLNG
GSCFCRDITKFKCTQYPLVHYALMFKLTTTCFAISLQSSYTGDDSYILEFFLPSSITDDQE
QDLLLLGSILVTMKEHFQSLRVASGVDFGEDDDKLSFEI IQALPDKKVHSHKIESIRVPFSG
FKSNATETMLIPQPVVQSSDPVNEKINVATVNGVVKEKKKTEKKRGKTEKTISLDVLQOY
FTGSLKDAAKSLGVCPTTMKRICRQHGISRWPSRKIKKVNRSITKLKRVIESVQGTGGGL
DLTSMVSSIPWTHGQTSAPLNSPNGSKPELPNTNNSPNHWSSDHSNPNPNSPELPP
SNGHKRSRTVDESAGTPTSHGSCDGNQLDEPKVPNQDPLFTVGGSPGLLFPYPYSRDHDVS
AASFAMPNRLGSLDHFRGMLIEDAGSSKDLRLNCPTAAFDDKFQDTNWMNNDNNSNNNL
YAPPKEEAIANVACEPSGSEMRTVTIKASYKDDIIRFRISGSGIMELKDEVAKRLKVDA
GTFDIKYLDLDDNEWVLIACDADLQECLEIPRSSRTKIVRLLVHDVTTNLGSSCESTGEL*

>G504 (69..1040)

CGTCGACCTCTTGACGATCATGAGACTGATTTTCGTGAAAATATCGTCATTATATCAAATT
AGAAGTTGATGGAACATGGGGGATTTCGAGCATAGGGCCGGCCATCCGCATCTCCCTC
CCGGGTTTCGGTTTCACCCGACTGATGAGGAAGTAGTAGTTTACCTCAAGAAGAAAG
CAGATTCTGTTCCACTTCCAGTCTCAATCATCGCAGAGATTGATCTTTACAAGTTTGATC
CTTGGGAGCTTCCAAGCAAGGCGAGTTTGGAGAGCACGAGTGGTACTTCTTTAGTCCTC
GGGATCGGAAGTATCCAAATGGGGTTAGGCCAAACCGGCGAGCAACTTCCGGTTATTGGA
AAGCAACGGGAACCGATAAACCGATATTTACGTGCAATAGTCACAAGTTGGTGTCAAGA
AAGCGCTTGTTTTTTACGGTGGAAAGCCTCCTAAAGGGATAAAAACAGATTGGATCATGC
ATGAATATCGCTCACTGATGGTAACCTTAGCACTGCGGCTAAGCCGCTGACTTAACCA
CGACAAGGAAAAAATCACTACGGCTAGACGATTGGGTTCTATGTAGGATCTATAAGAAGA
ATAGTTACACAAAGACCAACAATGGAGAGAGTATTACTTAGAGAGGATCTAATGGAAGGCA
TGCTCTCAAATCATCTGCTAATTCTTCTTCTACATCAGTACTAGACAACAACGACAACA
ATAATAACAATAACGAAGAACACTTTTTTCGACGGTATGGTCTGTTTCTTCAGACAAACGTT
CCTTGTGTGGTCAATACCGAATGGGCCACGAGGCCTCAGGATCATCTTCATTCCGATCTT
TCTTATCGAGCAAGAGGTTTCATCATAAGGTGATCTCAACAATGATAACTACAATGTCT
CTTTTGTTCGATGCTTAGTGAGATTCCTCAGAGTTCGGGGTTTCATGCAAATGGTGTTA
TGGATACGACGTCGTCTCTAGCTGATCATGGGGTTTTAAGACAGGCGTTTCAGCTTCTTA
ACATGAACCTGGCACTCATAATCTATATAGATATATATGTGTGTATCATATATGTATCTAT
GCAGGCCTAATATAGTTTACACATAAATCATCTGGGGCGGCCGCT

>G504 Amino Acid Sequence (domain in AA coordinates: TBD)

MENMGDSSIGPHPLPPGFRFHPDTDEELVVHYLKKKADSVLPVSIIEIDLYKFDPWE
LPSKASFGEHEWYFFSPRDRKYPNGVRPNRAATSGYWKATGTDKPIFTCNHKGKGVKAL
VFYGGKPPKGIKTDWIMHEYRLTDGNLSTAAKPPDLTTTRKNSLRLLDDVLCRIYKKNSS
QRPTMERVLLREDLMEGMLSKSSANSSSTSVLDNNDNNNNNNEEHFFDGMVVSDDKRLC
GQYRMGHEASGSSSFGSFLSSKRPHHTGDLNNDNYNVSVFVSMLEIPQSSGFHANGVMDT
TSSLADHGVLRLQAFQLPNMNWHS*

>G622 (248..2620)

TCTTTCTTTCTTCAATTGCGCGTCAAAATCTTCTCTTTCTTCTTCCCCCGCCGGTCCTTCA
CCAATCCTCTGATCTCTCTACACAGAACCTTTGATTTTGACCAACGTCGATGCATGTTT
ATGACTAGTCTCTTCTCAATCCTTCAATTTTCATCAATTCACGTCGATTTTCGTATCCGAT
TCGTTGTTCTAGCTCTTTGTGTGGTGTAGGGTTTTAAGATTTTGGAATTGGGGTTTGA
GTTTGTGATGTTTGAAGTCAAAATGGGGTCAAAGATGTGCATGAACGCTTCATGTGGTAC
GACTTCTACTGTTGAATGGAAGAAAGGTTGGCCTCTTCGATCTGGTCTTCTCGCTGATCT
CTGTTATCGTTGCGGATCTGCGTATGAGAGTTCTCTATTCTGTGAACAATTTCAAGGA
CCAATCTGGTTGGAGGGAATGCTATTTGTGTAGCAAGAGACTACATTGTGGATGCATTGC
TTCTAAGGTAACGATTGAGTTAATGGACTATGGTGGTGTGGTTGTAGTACATGTGCTTG
CTGCCATCAACTCAATTTGAACACAAGGGGTGAGAATCCAGGTGTTTTTAGCAGATTGCC
AATGAAAACGTTAGCTGATAGGCAACATGTAAATGGCGAAAGCGGAGGAAGAAACGAAGG
CGATCTCTTTTCTCAGCCACTAGTCATGGGCGGAGATAAAAGGGAAGAGTTTCATGCCCTCA
CCGTGGGTTTGGTAAGCTAATGAGTCCAGAAAGTACAACCACCGGGCATAGGCTGGATGC
TGATGGGGAAATGCATGAATCATCACCTTTACAGCCATCTTAAATATGGGGTTTGGCTGT
GAATCCGTTTACCCATCTTTTGAACCGAGGCTGTGCGAGGGAATGAAACACATCAGTCC
TTCTCAGTCCAACATGGTCCATTGCTCTGCTTCTAATATACTGCAAAAGCCATCAAGACC
TGCTATTTCAACTCCTCCTGTGGCTAGTAAATCCGCTCAGGCGCGGATTGGAAGGCCTCC
TGTCGAAGGGCGAGGGAGAGGCCACTTGCTTCCGCGGTATTGGCCAAAATATACGGATAA
AGAGGTTTCAGCAGATCTCTGGAAATTTGAATTTGAACATTGTACCTCTCTTTGAGAAAAC

TCTTAGTGCCAGTGATGCTGGTTCGCTAGTTCCTCCAAAAGCCTGTGCAGA
GGCATATTTTCTCCGATTAGTCAATCCGAAGGCATTCTTTGAAAAATCCAAGATGTGAG
GGGTAGGGAGTGGACGTTCCAGTTCAGATATTGGCCCAATAACAATAGTAGAATGTATGT
TTTAGAAGGTGTCACTCCATGCATACAGTCCATGATGCTACAGGCTGGTGATACAGTAAC
TTTCAGTCCGGTTGATCCTTGCGGAAAACATAATCATGGGTTCCAGGAAGGCAGCTAATGC
TGGAGACATGCAGGGTTGTGGGCTCACCAACGGAAACATCAACTGAGGACACATCATCGTC
TGGTGTAACAGAAAACCCACCCTCCATAAATGGTTCCTCGTGTATTTACTAATACCGAA
AGAGTTGAATGGTATGCCGTGAGAATTTGAACAGTGAGACTAACCGGGGCAGGATAGGTGA
TGATCCTACACGAGTTAAAGAGAAGAAGAGAACTCGAACCATTGGTGCAAAAAATAAGAG
ACTTCTTTTGCATAGTGAAGAATCTATGGAGCTGAGACTCACTTGGGAAGAAGCTCAGGA
CTTGCTTCGTCCCTCTCCTAGTGTAAGCCTACCATCGTTGTCTATTGAGGAGCAAGAAAT
TGAAGAATATGACGAACCTCCTGTCTTTGGAAAGAGGACTATAGTCACTACAAAACCTTC
AGGTGAACAGGAACGATGGGCAACTTGCGACGACTGCTCTAAATGGAGAAGGTTACCTGT
AGATGCTCTTCTTTCTTTTAAATGGACATGTATAGACAATGTTTGGGATGTGAGTAGGTG
TTCATGTTCTGCACCGGAGGAGAGTCTGAAGGAACCTGAGAATGTTCTTAAAGTAGGTAG
AGAGCACAAGAAGAGAAGAACTGGGGAAAGACAGGCAGCACAAAGTCAGCAAGAACCGTG
TGGTTTGGACGCACTGGCGAGTGCAGCAGTCTTAGGAGACACAATAGGCGAGCCAGAGGT
AGCGACCACGACCAGACATCCAAGGCACAGGGCTGGATGCTCTTGTCATCGTGTGCATTCA
GCCACCAAGTGGGAAAGGTAGGCACAAGCCTACATGTGGCTGCACTGTGTGTAGCACCCT
GAAGAGAAGGTTCAAGACGCTTATGATGAGGAGGAAGAAGAAGCAGTTGGAGCGCGATGT
AACAGCAGCAGAAGATAAGAAGAAGAACATGGAACCTGGCTGAGTCTGATAAGAGTAA
GGAGGAGAAGGAAGTAACACAGCGAGAATAGACCTGAACAGTGATCCATACAATAAAGA
AGATGTTGAAGCTGTTGCGGTGGAGAAAGAAGAGAGTCGAAAAAGAGCAATAGGACAGTG
TTCGGGCGTGGTGGCTCAAGACGCCAGTGATGTTTTAGGAGTTACAGAGTTAGAAGGAGA
GGGTAAGAAATGTTCTGTAAGAGCCGAGAGTTTCAAGCTGATATGGAAA

>G622 Amino Acid Sequence (domain in AA coordinates: TBD)

MFEVKMGSKMCMNASCGTSTSTVEWKKGWPLRSGLLADLCYRCGSAYESSLFCEQFHKDQS
GWRECYLCSKRLHCGCIASKVTIELMDYGGVGCSTCACCHQLNLNTRGENPGVFSRLPMK
TLADRQHVNGESGGRNEGDLFSQPLVMGGDKREEFMPHRGFGKLMSPESTTTGHRLLDAAG
EMHSSPLQPSLNMGLAVNPFSPSFATEAVEGMKHISPSQSNMVHCSASNILQKPSRPAI
STPPVASKSAQARIGRPPVEGRGRGHLPRYWPKYTDKEVQQISGNLNLNIVPLFEKTL
ASDAGRIGRLVLPKACAEAYFPPISSQSEGIPLKIQDVRGREWTFQFRYWPNNNSRMYVLE
GVTPCIQSMMLQAGDVTTFSRVDPGGKLIMGSRKAANAGDMQGCGLTNGTSTEDTSSSGV
TENPPSINGSSCISLIPKELNGMPENLNSETNGGRIGDDPTRVKEKKRTRTIGAKNKRLL
LHSEESMELRLTWEEAQDLLRSPSPVKPTIVVIEEQEIEEYDEPPVFGKRTIVTTKPSGE
QERWATCDDCSKWRRLPVDALLSFKWTCIDNVWDVSRSCSAPEESLKELENVLKVGREH
KKRRTGERQAAQSQQEPCGLDALASAAVLGDTIGEPEVATTTTRHPRHRAGCSCIVCIQPP
SGKGRHKPTCGCTVCSTVKRRFKTLMRRKKKQLERDVTAAEDKKKKDMELAESDKSKEE
KEVNTARIDLNSDPYNKEDVEAVAVEKEESRKRAIGQCSGVVAQDASDVLGVTELEGEK
NVREEPRVSS*

>G778 (50..1249)

TCTCAATAACACAAAACCTTTTAACTAGTAAAATACACAGATTTTAGGATGAGCCAATG
TGTTCCAAACTGTACATCGATGATACTCCGGCAGCAGCCACCACCACCGTCCGCTCCAC
CACAGCCGCAGACATCCCATATTAGACTACGAGGTAGCCGAGCTGACGTGGGAGAACGG
GCAACTAGGCTTGACACGGCTTAGGTCCACCGCAGTGACGGCTTCGTGACCAAGTACTC
CACAGGCGCCGGTGGAAACGTTGGAGTCGATAGTGGACCAAGCTACTCGCCTCCCTAACCC
TAAGCCCACGGATGAGCTCGTCCCGTGGTTCCATCATCGCTCCTCCAGGGCCGCGATGGC
AATGGACGCGCTTGTCCCTTGCTCCAACCTAGTACACGAGCAGCAGAGCAAGCCTGGTGG
CGTTGGCTCCACCCGGGTGGGGTCATGTAGCGATGGTCTGACCATGGGCGGTGGAAAACG
AGCAAGAGTGGCACCAGAGTGGAGCGGCGGGGAGTCAGCGGCTGACCATGGACACTTA
CGACGTAGGTTTCACTCAACATCAATGGGCTCGCACGATAACACAATCGACGATCATGA
CTCCGTCTGCCACAGCCGCCACAGATGGAGGACGAAGAAGAGAAGAAAGCCGAGGAAA
ATCATCAGTTTCAACCAAGAGAAGCAGAGCTGCTGTATTATTAACCAATCCGAAACGTAA
GAGGAGAGATAAAATCAATCAAAGGATGAAGACTTTGCAAAAACCTGGTTCCCAATTCAG
CAAGACGGATAAAGCATCTATGTTGGATGAAGTGATAGAGTATTTGAAGCAACTTCAAGC
ACAAGTGAGCATGATGAGCAGAATGAATATGCCTTCTATGATGCTTCTATGGCCATGCA
GCAACAACAACACTACAAATGTCTCTCATGTCCAATCCCATGGGTTTAGGGATGGGCAT

GTTCATGGAGTGAGGAGAGTAATGTAAATCTTTTAACTCGGTAGTAACAAGAGACAATG
TCTAAGTAGTGAATTCTCAATGTTTGTGTAAGTTTCTGCCTATGGAAGAGGCTTTCATT
TTTATGATTTTCACTATGTATGATCTCTCTTCACTGCATTTCTGGTTAGTAACGGCTTGT
CACCGATAAACTTTCTCGTTATGGAAAGTTAGAATAAAAAAAAAAAAAAAAAAAAAA

>G861 Amino Acid Sequence (domain in AA coordinates: 2-57)
MAREKIQIRKIDNATARQVTFSKRRRLFKKAEELSVLCDADVALIIFSSSTGKLFECSS
SMKEVLERHNLQSKNLEKLDQPSLELQLVENS DHARMSKEIADKSHRLRQMRGEELQGLD
IEELQQLBKALETGLTRVIETKSDKIMSEISELQKKGMQLMDENKRLRQQGTQLTEENER
LGMQICNNVHAHGAESENAAYVEGQSSSEITNAGNSTGAPVDSESSDTSRLRLPLPYGG

*
>G938 (1..1755)
ATGATGATGTTTAAACGAGATGGGAATGTATGGAAACATGGATTTCTTCTCTCCTCCACA
TCTCTCGATGTGTGTCCATTACCACAAGCTGAACAAGAACCTGTAGTTGAAGATGTTCGAC
TACACCGATGATGAGATGGATGTGGATGAGCTTGAGAAGAGGATGTGGAGAGACAAAATG
CGTTTGAAACGCTCTCAAGGAGCAACAGAGTAAGTGTAAAGAAGGCGTCGATGGTTCGAAA
CAGAGGCAGTCGCAAGAGCAAGCTAGGAGGAAGAAAATGTCTAGAGCCCAAGATGGGATC
TTGAAGTATATGTTGAAGATGATGGAAGTTTGTAAAGCTCAAGGCTTTGTTTATGGTATT
ATTCTTGAGAAGGGTAAGCCTGTGACTGGTGCCTTCGGATAATTTGAGGGAATGGTGGAAA
GATAAGGTTAGGTTTGATCGTAATGGTCCAGCTGCTATTGCTAAGTATCAGTCAGAGAAT
AATATTTCTGGAGGGAGTAATGATTGTAACAGCTTGGTTGGTCCAACACCGCATACGCTT
CAGGAGCTTTCAGGACACGACTCTTGGTTCGCTTTTATCGGCTTTGATGCAACATTGTGAT
CCACCGCAGAGACGGTTTCTCTTGGAGAAAGGAGTTTCTCCACCTTGGTGGCCTAATGGG
AATGAAGAGTGGTGGCCTCAGCTTGGTTTACCAAATGAGCAAGGTCCTCCTCCTTATAAG
AAGCCTCATGATTTGAAGAAAGCTTGGAAAGTCGGTGTTTTAACTGCGGTGATCAAGCAT
ATGTCGCCCGATATTGCGAAGATCCGTAAGCTTGTGAGGCAATCAAAATGCTTGCAGGAT
AAGATGACGGCGAAAGAGAGTGCTACTTGGCTTGCCATTATTAACCAAGAAGAGGTTGTG
GCTCGGGAGCTTTATCCCGAGTCATGCCCTCCTCTTCTTCTTCTCATCATTAGGAAGC
GGGTCGCTTCTCATTATGATTGTAGCGAGTATGACGTTGAAGGTTTCGAGAAGGAACAA
CATGGTTTCGATGTGGAAGAGCGGAACAGAGATAGTGATGATGCATCCTCTAGCAAGC
TTTGGGGTTGCTAAATGCAACATTTTCCATAAAGGAGGAGGTCGCCACCACGGTAAAC
TTAGAGTTCACGAGAAAGAGGAAGCAGAACAAATGATATGAATGTTATGGTAATGGACAGA
TCAGCAGGTTTACACTTGTGAGAAATGGTCAGTGTCTCACAGCAAAATGAATCTTGGATTT
CAAGACAGGAGTTCAAGGGACAACACCAGATGGTTTGTCCATATAGAGACAATCGTTTA
GCGTATGGAGCATCCAAGTTTCTATATGGGTGGAATGAACTAGTAGTTCCTCAGCAACCA
GTCCAACCGATCGACCTATCGGGCGTTGGAGTTCGGGAAAACGGGCAGAAAGATGATCACC
GAGCTTATGGCCATGACGACAGAAATGTCCAAAGCAACCAACGCCCTCTACTTTGATG
GAAACCAAAAGCATGGTCATTGATGCAAAAGCAGCTCAGAATCAGCAGTGAATTTCAAC
AGTGGCAATCAAATGTTTATGCAACAAGGGACGAACAACGGGGTTAACAATCGGTTCCAG
ATGGTGTTTGATTTCGACACCATTCGATATGGCAGCATTCGATTACAGAGATGATTGGCAA
ACCGGAGCAATGGAAGGAATGGGGAAGCAGCAGCAGCAGCAGCAGCAGCAAGATGTA
TCAATATGGTTCTGA

>G938 Amino Acid Sequence (domain in AA coordinates: 96-104)
MMMFNEMGMYGNMDFSSSTSLDVCPLPQAEQEPVVEDVDYTDDEMVDLEKRMWRDKM
RLKRLKEQQSKCKEAGVDSKQSQSQEQARRKKMSRAQDGLKYLKMMEVCKAQGFVYGI
IPEKKGKPVGTASDNLREWWKDKVRFDRNGPAAIAKYQSENNISGGSNDCNSLVGPTPHTL
QELQD'TT'LGSLLSALMQHCDPPQRRFPLEKGVSPPPWWPNGNEEWWPQLGLPNEQGPPPYK
KPHDLKKAWKVGVLTAIVKHMSPDIAKIRKLVRQSKCLQDKMTAKESATWLAIINQEEVV
ARELYPESCPPLSSSSSLGSGSLINDCSEYDVEGFEKEQHGFDVEERKPEIVMMHPLAS
FGVAKMQHFPIKEEVATTNLEFTRKRKQNNDMNVMVMDRSAGYTCENGQCPHSMNLF
QDRSSRDNHQMVCPYRDNRLAYGASKFHMGMKLVVPQQPVQPIDLSGVGPENGQKMIT
ELMAMYDRNVQSNQTPPTLMENQSMVIDAKAAQNQQLNFNSGNQMFQQGTNNGVNNRFQ
MVFDSTPFDMAAFDYRDDWQTGAMEGMGKQQQQQQQQQDVSIWF*

>G965 (73..1956)
GATTCTCTGTATGCTGAATCCTTACAGGATCCAAGAGCTTTGGAAAAAGATATAAT
GAATAACAAGATATGGGTTTAGCTACTACAAC'TCTTCTATGTACAAAGATTATCATCAT
CACCAAGGAATCTTTTCTTCTCTAATGGATTCCACCGATCATCATCAACCACTCATCAG
GAGGAAGTAGATGAATCCGCCGTCGTCTCCGGTGCTCAAATTCGGTTTATGAAACCGCC

GGAATGTTGTCTGAAATGTTTGCTTACCCTGGCGGAGGTGGCGGCGGTTCCGGTGGAGAG
ATTCTTGATCAGTCTACTAAACAGTTGCTAGAGCAACAAAACCGTCACAACAACAACAT
AACTCAACTCTTCATATGTTATTACCAAATCATCATCAAGGTTTTGCTTTCACCGACGAA
AACACTATGCGAGCCGAGCAACAACAACACTTTACATGGCCATCTTCCTCCTCCGATCAT
CATCAAAACCGAGATATGATCGGAACCGTCCACGTGGAAGGAGGAAAGGGTTTTGTCTTTA
TCTCTCTCATCTTCATTAGCCGAGCTAAAGCCGAGGAATATAGAAGCATTTATTGTGCA
GCCGTTGATGGAACCTCTTCTTCTTCTAACGCATCCGCTCATCATCATCAATTCAATCAG
TTCAAGAATCTTCTTCTTGAGAATTCTTCTTCTCAACATCATCACCATCAAGTTGTTGGA
CATTTTGGTTTCATCATCATCATCTCCCATGGCGGCTTCTTCATCCATTGGAGGGATCTAC
ACGTTGAGGAATTCGAAATATACGAAACCGGCTCAAGAGTTGTTGGAAGAGTTTTGTAGT
GTTGGAAGAGGACATTTCAAGAAGAACAACCTTAGTAGGAACAACCTCAAACCCTAATACT
ACCGGTGGAGGAGGAGGCGGAGGGTCTCGTCATCGGCCGGAACAGCTAATGATAGTCCT
CCTTGTCTCCGGCTGATCGGATTGAACATCAAAGAAGAAAAGTCAAGCTACTATCTATG
CTTGAAGAGGTGGACCGGATGACCAACTACTGCGAACAATGCAATGGTAGTGAAC
TCATTGACCAAGTAATGGGTACGGCGCGGCGGTTCCGTACACGACATTAGCTCAAAAG
GCAATGTCTAGGCATTTCCGGTGTGTTGAAAGACGCGGTAGCGTTTCAGCTTAAACGCAGC
TGTGAGCTTCTAGGGGATAAAGAGGCGGAGGGGCTGCATCCTCGGGGTTAACCAGAGG
GAAACGCCGCGATTGCGTTTTGCTAGAGCAGAGTTTGCCTCAGCAACGAGCGTTTCATCAT
ATGGGTATGATGGAGCAAGAGGCATGGAGACCGCAACGTGGTTTTGCCGTAACGCTCCGTT
AATATCCTTAGAGCTTGGCTATTCGAGCATTTTCTTAATCCGTACCAAGCGATGCTGAT
AAGCACCTCTTAGCACGACAGACTGGTTTTATCCAGAAATCAGGTGTCAAATGGTTTCATA
AATGCTAGGGTTTCGCTATGGAACCAATGGTGGAGAGATGTATCAACAAGAAGCAAAA
GAAAGAGAAGAAGCAGAAGAAGAAAATGAAAATCAACAACAACAAGAAGACAGCAACAA
ACAAACAACAACGACACGAAACCAACAACAATGAAAACAACCTTCACTGTCATAACCGCA
CAAACCTCAACGACGATGACATCGACACATCAGGAAACGACTCTTCATTCTCTCTCTCC
GTCGCGCCGCTTCTCAGCGCGGTTTCAGACGCGTTACCGTCCGACGTCAGCAAGAC
GTCAGTGACTTCCACGTCGACGAGATGGTGTGAACGTCATAAGATTGGGACCAACAG
ACTGGTGACGTGTCTCTTACGCTTGGTCTACGCCACTCTGGCAATATTCCTGATAAGAAC
ACTCTTTCTCCGTTAGAGACTTTGGAGATTTTTAGTCTTCTTTGTTTCTCAATTTATTC
ATC

>G965 Amino Acid Sequence (domain in AA coordinates: 423-486)
MGLATTTSSMSQDYHHHQIFSFNSGFHRSSSTTHQEEVDES AVVSGAQIPVYETAGMLS
EMFAYPGGGGGSGGEILDQSTKQLLEQQNRHNNNNSTLHMLLPNHHQGFATDENTMQ
PQQQQHFTWPSSSSDHQNDRDMIGTVHVEGGKGLSLSLSSSLAAAKAEFYRSIYCAAVDG
TSSSSNASAHHQFNQFKNLLLENNSSQHQQVVGHFSSSSSPMASSSIGGIYTLRN
SKYTKPAQELLEEFCSVGRGHFKKNLSRNNSNPNTTGGGGGGSSSSAGTANDSPPLSP
ADRIEHQRRKVKLLSMLEEVDRRYNHYCEQMOMVVSFDQVMGYGAAPYTTLAQKAMSR
HFRCLKDAVAVQLKRSCELLGDKEAAGAASSGLTKGETPRLRLLEQSLRQQR AFHMGMM
EQEAWRPQRGLPERSVNILRAWLFEHFLNPYPSDADKHLARQTGLSRNQVSNWF INARV
RLWKPMVBEMYQQEAKEREEAEEENENQQQRRQQQTNNNDTKPNNNENNFVITAQTPT
TMTSTHHENDSSFLSSVAAASHGGSDAFTVATCQQQDVSDPHVDGDGVNVIRFGTKQTDV
SLTLGLRHSNIPDKNTSFSVRDFGDF*

>G1143 (54..677)

AAATAAGAATATAAACACTTTTGTCTGAAAAATTATCAAAGAAGAAGAAATAAATGGGTG
GAGGAAGCAGATTTCAAGAACCAGTGAGGATGAGCCGTAGGAAACAAGTAACAAAAGAGA
AGGAAGAAGATGAAAACTTCAAATCTCCAAATCTTGAAGCAGAGAGACGTAGAAGAGAGA
AGCTTCATTGTGCGCTTATGGCTCTGCGATCTCATGTCCCCATTGTCACCAACATGACTA
AAGCAAGTATTGTTGAAGATGCGATTACTTACATAGGAGAGCTTCAAACAATGTTAAGA
ATCTCTTAGAGACATTTTCATGAAATGGAAGAAGCTCCTCTGAGATTGATGAAGAACAAA
CGGATCCAAATGATAAAACCTGAAGTTGAAACTAGTGATCTTAACGAAGAGATGAAGAAAC
TCGGAATCGAGGAGAATGTGCAATTGTGTAAGATTGGGGAGAGGAAGTTTTGGTTAAAGA
TCATAACAGAGAAGAGAGATGGGATCTTTACTAAATTCATGGAGGTTATGAGATTTCTCG
GATTCGAGATTATCGATATTAGTCTAACAACTTCAAATGGAGCAATTCTTATTAGTGCCT
CTGTTTCAGACACAGAACTCTGTGATGTTGAACAGACAAAAGATTTTCTTTTGGAAAGTTA
TGAGAAGCAATCCATAAGTATTAATTATATACATCTTGGAATTTCTTGATCTAATAACA
TTTCCATTGGTTTTTATTACATTGTTGTTCCATTTTAAATATGATATGATTGAGATGAAA
AAGAGTTTGTGTTACAAGCCAATGA

>G1143 Amino Acid Sequence (domain in AA coordinates:33-82)
MGGGSRFQEPVRMSRRKQVTKEKEEDENFKSPNLEABRRRREKLHCRMLALRSHVPIVTN
MTKASIVEDAITTYIGELQNNVKNLLET FHEMEEAPPEIDEEQTDPMIKPEVETSDLNEM
KKLGIEENVQLCKIGERKFWLKIITEKRDGIFTKFMEVMRFLGFEIIDISLTTSNGAILI
SASVQTQELCDVEQTKDFLLEVMSNP*

>G1190 (209..2020)

TCCTGTCCCAAAACAAAAGACTTGAGAGTGTGTCTTTAGAGAGAGATCTTCTCTCTTT
ATCTTACGACTCTCACTTCTTATCTCAAATCTACTTCAACTCTATTTCCAGTCTCCACAT
TTTCCACAAAATTTCAACTCTTGTCTCTTCTCCTCCAAAGTAAAAACAAATCGTTGCAAG
TGAGGTTTGGTTTGGTGTATAGAAATTATGAAGAGCGGGAAGCAATCTTCGCAACCTGA
AAAGGGTACTTCCAGGATCTTGTCACTGACTGTCTGTTTATCGCATTTTGGGTTTCTC
CTTCTACCTCGGTGGTATATTTTGCTCTGAGAGAGACAAGATTGTAGCCAAGGATGTAC
AAGGACGACTACAAAGGCTGTAGCTTCCCTAAAGAACCCTACAGCTACTCCTATTCAAAT
CAAATCCGTTTCTTTCCCGGAGTGCAGGTCAGAGTTCCAAGATTACACCCCGTGACCCGA
TCCAAAGAGGTGGAAGAAGTATGGTGTCCATCGCTTAAGTTTCTTGGAGCGTCATTGTCC
TCCGGTATATGAAAAGAATGAGTGTGATTCCACCACCAGACGGGTATAAACCCTAT
AAGATGGCCCCAAGAGCCGAGAACAGTGTGGTACAGGAACGTGCCTTATGATTGGATCAA
TAAGCAAAAGTCTAACCAGCATTGGCTTAAGAAAGAAGGAGATAAGTTCCATTTCCCTGG
TGGTGGTACCATGTTCCCTCGTGGAGTTAGTCACTATGTTGATTGTGCAAGATCTGAT
TCCTGAAATGAAAGACGGAACAGTCAGGACCGCCATTGATACTGGCTGTGGGGTTGCGAG
CTGGGGAGGCGATCTTTTGGACCGTGGGATACTATCACTCTCTTGTCTCAAGAGATAA
CCATGAAGCTCAGGTTCAATTTGCTCTTGAACGTGGAATTCCTGCGATTCTCGGGATCAT
CTCTACGCAACGTCTCCCTTTTCTTCAAATGCATTTGATATGGCTCATTGTTCAAGATG
TCTTATTTCCCTGGACAGAATTTGGTGGAACTTATTTACTTGAGATTACCGTATAGTTCCG
ACCTGGAGGTTTTTGGGTTCTTTCTGGTCCACCTGTGAACATAATAGACGATGGCGTGG
ATGGAACACAACCATGGAAGATCAGAAATCTGACTACAACAAGCTTCAGTCAC'TTCTAAC
CTCCATGTGTTTCAAAAAGTACGCTCAAAAAGATGACATAGCCGTGTGGCAGAACTCTC
AGACAAATCTTGCTATGACAAAATCGCTAAGAACATGGAAGCTTACCCTCCCAAATGTGA
CGACAGTATAGAACCTGATTCTGCTTGGTACACTCCACTCCGTCTTGGCTGGTTGCCCC
GACACCTAAAGTCAAGAAGTCTGGTCTCGGATCAATCCAAAATGGCCCCGAGAGGTTACA
TGTCGCGCCCCGAGAGAATCGGTGATGTTACCGGAGGGAGTGCAGAACAGTTTGAAACACGA
TGATGGTAAATGGAAGAACAGAGTTAAGCATTACAAGAAAGTTTTACCAGCTCTTGGGAC
AGACAAGATAAGAAATGTTATGGATATGAACACTGTTTATGGAGGTTTCTCTGCGGCCCT
CATTGAGGATCCCATTTGGGTCATGAACGTTGTATCATCGTACAGCGCAAATTCGCTTCC
TGTTGTCTTTGATCGCGTCTCATCGGGACTTACCACGACTGGTGCGAAGCTTTCTCAAC
GTATCCAAGAACATATGATCTTCTTCACTCGACAGTCTTTTACCTTGGAGAGTCACAG
GTGTGAGATGAAGTACATTTTGCTAGAGATGGACAGGATCTTGGCGCCGAGTGGATATGT
TATAATCCGAGAATCGAGTTATTTTATGACGCAATCAACGTTAGCGAAAGGGATAAG
GTGGAGTTGCCGGAGAGAGGAGACTGAGTATGCAGTCAAAGTGAGAAGATTCTGGTTTG
CCAGAAAAAGCTATGGTTTTTCGTCAAACCAACCTCTTGTATGAGACCACCTGTATCATAG
TGTTTATCATCTCCTGTGATGCACACTACAGAGAGAAGGATCTAGTCCTTTGAGTCCAAG
ATATAGCTCTATAAACAATCTCCTTTTTTTGTTCTCTTTAATTTCTTGGGTATTTACGG
TATAGATTGATATTATATATTTTTTAATTATATTTTAATATATAGATATATTAGTATGT
GGTTTAAACACTATTATTATCAAGGTCTTAAAGATTGCTTTGCAAGAGTTAAAAAATGT
TGGAGTAAGGACCTCTTGATTAATAAATTGACTGACGCAGCAA
>G1190 Amino Acid Sequence (domain in AA coordinates: entire protein)
MKSGKQSSQPEKGTSLRLSLTVLFIAFCGFSFYLGIGFCSEKD KIVAKDVTRTTTKAVAS
PKEPTATPIQIKSVSFPECGSEFQDYTPCTDPKRWKYGVHRLSFLERHCPPVYEKNECL
IPPPDGYKPPIRWPKSREQCWYRNPYPYDWINKQKSNQHWLKEGDKFHPGGGTMFPRGV
SHYVDLMQDLIPEMKDGTVRTAIDTGCGVASWGGDLLDRGILSLSLAPRDNHEAQVFAL
ERGIPAILGIISTQRLPFPNADFMAHCSRCLIPWTEFGGIYLLLEIHRIVRPGGFVWL
PPVYNRRWRGWNTTMDQKSDYNKLQSLTSMCFKKAQKDDIAVWQKLSKSCYDKIA
KNMEAYPPKDDSI EPDSAWYTPLRPCVVAPTPKVKSGLSIPKWPERLHVAPERIGDV
HGGSANSLKHDDGKWKNRVKHYKKVLPALGTDKIRNVMDMNTVYGGFSAALIEDPIWVMN
VVSSYSANSLPVVFDRLIGTYHDWCEAFSTYPRTYDLHLDSLFTLESHRCMKYILLE
MDRILRPSGYV IRESSYFMDAITTLAKGIRWSCRRETEYAVKSEKILVCQKKLWFSSN
QTS*

>G1198 (230..1675)

TCCTTTTCAAATTCCAATCATTGATCAACTAATCAAGAATTAATTATAAGACTTTGCAAT
CTCTCTCCCTCTCCCTCTCCCTAGCTAGTTCTCTCTTGTGTTTCTTAAGCTCGAGCTTCTC
TCAATAGTGATTATCATCTTTTTCATCATTTCAGATTAAATGTGTTTTGCAGAAAAGAG
ACTAATCAAGAAGAGATATCATCAATTGAAGCTGTTTTCTTGAGTAGAGATGGCGAACCA
TAGAATGAGCGAAGCTACAAACCATAACCACAATCATCATCTTCCTTATTCACTTATTCA
TGGTCTCAACAACAATCATCCATCTTCTGGTTTCATTAACCAAGATGGATCGTCCAGTTT
CGATTTTGGAGAGCTAGAAGAAGCAATTGTTCTGCAAGGTGTCAAGTATAGGAACGAGGA
AGCCAAGCCACCTTTATTAGGAGGAGGAGGAGGAGCTACGACTCTGGAGATGTTCCCTTC
GTGGCCAATCAGAACTCACCAAACTCTTCCTACTGAGAGTTCCAAGTCAGGAGGAGAGAG
CAGCGATTTCAGGATCGGCTAATTTCTCCGGCAAAGCTGAAAGTCAACAACCGGAGTCTCC
TAGTAGTAGCAAAACATCATCTCATGCTTCAACCTCATCATAATAACATGGCAAACTCAAG
TTCAACATCTGGACTTCCTTCCACTTCTCGAACTTTAGCTCCTCTAAACCTTCGGAAGA
TAAGAGGAAGGCTACAACCTCAGGCAAACAGCTTGATGCTAAGACGTTGAGACGTTTGGC
CCAAAATAGAGAAGCTGCTCGCAAAAGCCGTCTTAGGAAAAAGGCGTATGTGCAACAGCT
AGAATCAAGTAGGATAAAGCTTTCCCAATTGGAGCAAGAACTTCAGCGAGCTCGTTCTCA
GGGGCTGTTCATGGGTGGTTGTGGACCACCAGGACCTAACATCACTTCGGGAGCTGCAAT
ATTTGACATGGAATATGGGAGATGGCTAGAGGATGATAACCGGCATATGTGCGGAGATTCTG
AACCCTGCTTCAGGCTCATTTATCTGACAATGATTTAAGGTTGATCGTTGACGGTTACAT
TGCTCATTTTGTGATGAGATATTCCGATTAAAAGCCGTGGCAGCGAAAGCCGATGTTTTTCA
CCTCATCATTTGGGACATGGATGTCCCCAGCCGAACGTTGTTTTATTGGATGGCTGGTTT
CCGTCCATCCGACCTAATCAAGATATTGGTGTGCAAAATGGATCTATTGACGGAGCAACA
ACTGATGGGAATATATAGCCTACAACACTCGTCGCAACAAGCAGAGGAGGCTCTCTCGCA
AGGCCTCGAACAACCTTCAGCAATCTCTCATCGATACTCTCGCCGCATCTCCAGTCATTGA
CGGAATGCAACAATGGCTGTGCTCTCGGAAAGATCTCTAATCTCGAAGGCTTTATCCG
CCAGGCTGATAACTTGAGGCAGCAGACCGTTCACCAGCTGAGGCGGATCTTGACCGTCCG
ACAAGCTGCACGGTGTTCCTAGTCATCGGAGAGTACTATGGACGGCTCAGAGCTCTTAG
CTCCCTTTGGTTGTGTCAGCCCCAGAGAGACACTGATGAGTGATGAAACCTCTTGTCAAAC
GACGACGGATTGTGAGATTTGTCAGTCATCTCGGAACCACTTCTCCAATTTCTGAATGGA
ATGAACTTTGTATAACTAAAAGGCCAAGTTTCATTGTCTGTGTAATTTACCTATTTT
CTTTAAAGTTGTACTAGAGAAAAGATAGGATCTTCCTTCG

>G1198 Amino Acid Sequence (domain in AA coordinates: 173-223)

MANHRMSEATNHNHNNHLPYSLIHGLNNHNPSSGFINQDSSSFDFGELEEAIVLQGVKY
RNEEAKPPLLGGGGGATTLEMFPSWPIRTHQTLPTSSKSGGESSDSGSANFSGKAESQO
PESPMSSKHLMQLPHNNMANSSTSLPSTSRTLAPPKPSSEDKRKATTSQKQLDAKTL
RRLAQNREAAARKSRLRKAYVQQLSSRIKLSQLEQELQARSQGLFMGGCGPPGPNITS
GAAIFDMEYGRWLEDNRHMSIIRGLQAHLSNDLRLIVDGYIAHFDEIFRLKAVAACA
DVFHLIIGTWMSPAERCFIWMAGFRPSDLIKILVSQMDLLTEQQLMGIYSLQHSSQQAEE
ALSQGLEQLQQLIDTLAASPVIDGMQOMAVALGKISNLEGFIRQADNLRQQTVHQLRRI
LTVRQAARCFIVIGEYYGRLRALSSLWLSRPRETLMSDETSCQTITDLQIVQSSRNHFSN
F*

>G1226 (212..1159)

CTGCATTTATTAAGAACAGTTTAGAAAGTGTCAACCCCTAAAGGAATGTTTTAGTTTAG
AGGAAAGAGAGAGAAGAAGCAGCAGCAGAGAAGTTGTTAATTTGAAGACTATTTGAGGA
AAGACACCTATATCTAAATACTCAAAGTTACAAAATATTACTTCAGAAAACAGTTCCAT
TAGAGAGACTCATAAGCTTCTCATCTAATTATGAGTGGATTGATGAGTTTGGTGAATT
AGAAGACCAATTTGGTCAGATTTCAGACACTACTATGGAAGAGAAGATACCATTTCTGCA
AATGCTTCAATGCATAGAACACCCTTTTACAACAACAGAACCAATCAGTTTCTCCAATC
ACTTCTCCAGATCCAAACCCTAGAATCAAAGAGCTGTCTACCCCTTGAAACAAACATCAA
AAGAGATCCGGGTCAAACAGATGACCCGGAAGGATCCAAGAACAGAAAACGGAGCAGT
AACGGTCAAAGAAAAAAGAAAACGGAACGTACAAGAGCTCAAAGAACAAGACGAAGT
TGAAAACCAAAGGATGACTACATTTGCCGTGCAACGTAATCGAAGACGACAAATGAACGA
ACACTTAAACTCTCTCCGATCTCTCATGCCTCCTTCGTTTCTTCAACGGGGTGACCAAGC
TTCGAFTGTAGGAGGGGCAATAGATTTTCATCAAGGAACAGAGCAACTCTTGCAATCTCT
AGAAGCTGAGAAACGAAAGGATGGAACGTGATGAAACTCCTAAAACGGCGTCGTGTTCTTC
ATCTTCGTCTCTTGATGCACTAACTCTTCTATTCTAGCGTGTCTACGACGTGCGAAAA
TGGATTTACGGCGAGATTTCGGCGGTGGAGATACGACAGAAGTGGAGGCTACGGTGATACA

GAACCATGTGAGCTTAAAAGTTCGGTGTAAAGAGAGGAAAAACGACAGATCTTAAAAGCTAT
TGTCTCGATTGAAGAACTAAAGCTTGCGATTCTACATCTCACTATCTCTTCTTCTTGA
CTTTGTCACTACTCTTTCAATCTCAAGATGGAAGATGGTTGTAAATTAGGATCAGCAGA
TGAGATAGCGACAGCCGTTTCATCAGATCTTCGAGCAAATCAACGGTGAAGTCATGTGGTC
AAATCTTAGTTCGAACTTAGTTGACTTTTGACTCCTAGTAACGTGTGTAAACTTTAGGTTA
CAAAGAAAAGGGACGTGATATAAATAAGAAAAACCAAAGAGGTGAAATTTTGGGAGTTTT
AATTATTATCTTATACTTTTTGGATTTTAGATTAGTAGCAAACCTCGCAGTGTCTACGAT
GACATTATTATTGGTCACATGAAGGTTTAGGTTAAAAA

>G1226 Amino Acid Sequence (domain in AA coordinates:115-174)
MSG LMSFGELEDQFGQISD TTMEEKIPFLQMLQCIEHPFTTTEPNQFLQSL LQIQTLESK
SCLTLETNIKRDPGQTD DPEKDPRTENGAVTVKEKRKRKRTRAPKNKDEVENQRMTHIAV
ERNRRRQMN EHLNLSRLMPPSFLQRGDQASIVGG AIDFIKELEQLLSLEAEKRKDGTD
ETPKTASCS SSSSLACTNSSISSVSTTSENGFTARFGGGDTTEVEATVIQNHVSLKVRCK
RGKRQILKAIVSIEELKLAILHLTISSSFDFVIYSFNLKMEDGCKLGSAD EIAATAVHQIF
EQINGEVMWSNLSRT*

>G1451 (124..2559)

TTTGTACTTCCGGAGCTAAAGAGTTATAGCTACTGTAGTAGCTGGAAGTGAAGAAGATTT
TTTAATAGATTGTACGGAAAAATTAGGGTTTTCAAAGTTTGGTTTTCTTGAAGTTGAATTA
GACATGAAGCTGTCAACATCTGGATTGGGTCAACAGGGTCATGAAGGAGAGAAGTGTCTG
AATTCTGAGCTATGGCATGCTTGTGCTGGACCATTAGTCTCTCTTCCATCATCTGGTAGT
CGAGTTGTTTACTTTCCACAGGGTCACAGTGAACAGGTAGCTGCTACAACCTAATAAGGAA
GTTGATGGTCACATAACCAATTACCCAAGCCTACCACCACAATTGATATGCCAGCTCCAT
AATGTTACAATGCATGCAGATGTTGAGACGGATGAAGTCTATGCTCAAATGACACTTCAA
CCATTGACACCGGAGGAGCAGAAGGAAACATTTGTACCGATTGAGTTGGGGATACCGAGT
AAGCAACCTAGTAATTATTTTTGTAAGACTCTCACAGCTAGTGATACCACTACACATGGA
GGGTTTTCTGTTCTTAGACGTGCTGCTGAGAAAGTGTTTCCCTCCATTGGATTACACACTG
CAGCCACCAGCTCAAGAACTGATTGCAAGGGATCTCCATGATGTTGAATGGAAGTTTAGG
CATATCTTTTCGGGGACAGCCCAAACGGCATCTCCTAACTACTGGATGGAGTGTCTTTGTC
AGTGCCAAAGCAGCTAGTAGTGGAGATTCTGTCTATTTTCATCAGGAATGAAAAGAATCAA
CTCTTTTTGGGAATTTCGTCTATGCCACTCGGCCCGCAGACTATTGTACCATCATCTGTTTTA
TCTAGTGATAGCATGCATATTGGACTCCTTGCTGCTGCTGCACATGCTTCTGCAACTAAT
AGCTGTTTCACTGTTTTCTTTTCATCCAAGGGCTAGCCAATCTGAGTTTGTGATACAACCT
TCCAAGTACATTAAAGCCGTTTTTTCACACGCGTATTTTCAGTTGGGATGCGCTTTTCGCATG
CTCTTCGAGACAGAAGAGTTCGAGTGTCCGCAGGTACATGGGTACTATAACTGGTATTAGT
GATCTAGATTCTGTTGCTTGGCCAACTCTCATTGGCGATCTGTGAAGGTTGGTTGGGAT
GAATCGACTGCAGGGGAGAGACAGCCAAGGGTTTCTTTATGGGAGATTGAGCCTCTGACT
ACCTTTCTTATGTATCCATCTCTTTTCTCTCAGACTAAAACGTCCATGGCATGCTGGC
ACATCATCTTTTGCCTGATGGAAGGGGTGATTTGGGAAGTGGTCTAACATGGCTAAGAGGG
GGAGGTGGAGAGCAGCAAGGTTTGCTTCTCTAAATTATCCATCTGTTGGTTTGTTCCTCA
TGGATGCAACAAAGGCTGGATCTCAGTCAAATGGGGACTGATAATAATCAGCAATACCAA
GCAATGTTAGCTGCTGGGTTGCAGAACATCGGCGGTGGAGATCCTTTAAGACAGCAGTTT
GTACAGCTGCAAGAGCCTCACCACCAATATCTTCAACAATCAGCTTCCCATAAATTCTGAT
TTGATGCTTCAGCAGCAACAGCAGCAACAGCGTCACGCCATCTCATGCATGCTCAAACA
CAGATTATGAGTGAGAATCTTCCGACAGCAGAATATGCGACAAGAAGTTAGTAACCAACCA
GCTGGACAGCAGCAACAGCTACAGCAACCGGACCAAATGCATATCTTAATGCTTTCAAA
ATGCAAAATGGCCATCTTCAACAGTGGCAGCAGCAATCAGAGATGCCATCTCCCTCGTTC
ATGAAGTCAGATTTTACTGACTCAAGCAACAAATTTGCAACAACCTGCTAGTCCGGCTTCT
GGAGATGGCAATCTTTTGAATTTTTCTATAACCGGTCACTCTGTACTCCCTGAGCAGTTA
ACAACAGAGGGTGGTCTCCAAAGCATCCAACACTTTTTCTGAACCGTTGTCACTTCCA
CAAGCCTATCCTGGGAAGAGTCTTGCTCTAGAACCCGGAATCCGACAGATCCCTCTCTT
TTCGGTGTGATCCCGACTCTGGACTCTTCTCTCCCGAGTACGGTTCCCGCTTTGCTTCT
TCATCAGGAGATGCTGAAGCTTCCCTATGTCACTAACAGATTGAGGATTTTCAAGATTC
TTATATAGCTGCATGCAAGACACAACCTCATGAGTTATTGCATGGAGCTGGACAGATTAAC
TCGTCCAACCAACCAAGAACCTTTGTAAAGGTTTATAAATCTGGTTTCGGTTGGGCGTTCA
TTAGACATCTCCCGATTGAGCAGCTACCACGAGCTGCGAGAAGAGTTAGGGAAGATGTTT
GCTATCGAAGGGTGTGTTGGAAGACCCCTTAGATCAGGCTGGCAGCTTGTATTCTGTTGAC
AAGGAAAATGATATTCTTCTCTCTTGGTGATGACCCATGGGAGTCATTTGTGAATAACGTT

TGGTACATAAAGATACTATCACCAGAAGATGTGCATCAAATGGGAGATCATGGAGAAGGC
AGTGGTGGGTTATTCCTCGCAAACCCGACCCATCTCTAGAAGCTGCTTCGGTGTAGTCT
CATCATGCTACAACGCGGGAGCCCTTTGTTTCCATTGAAGTCGTTTCCACTCATCTTT
ATATGCCATTGCTTCGCATCTCTCTCGTTTTGACGTTTTAGAAAGAAACATAATCATAT
TTGTGAGTATGGGTCTGAAACTTTAGGACGTACTTTAGCTTGTATTAGACAGACACTCT
CGTCATAAACATAAGAACCTTTATGTAGCTGTCTCAGGGTAACTAACTTTTCTAG

>G1451 Amino Acid Sequence (domain in AA coordinates: 22-357)

MKLSTSGLGQQGHEGEKCLNSELWHACAGPLVSLPSSGSRVVYFPQGHSEQVAATTNKEV
DGHIPNYPSPQLICQLHNVMTHADVETDEVYAQMTLQPLTPPEQKETFPVPIELGIPSK
QPSNYFCKTLTASDTSTHGGFSVPRRAAEKVFPPLDYTLQPPAQELIARDLHDVIEWKFRH
IFRGQPKRHLLTTGWSVFVSAKRLVAGDSVIFIRNEKNQLFLGIRHATRPQTIVPSSVLS
SDSMHIGLLAAAAHASATNSCFTVFFHPRASQSEFVIQLSKYIKAVFHTRISVGMFRFML
FETEISSVRRYMGTTITGISDLSVRWPNSHWRSVKVGWDESTAGERQPRVSLWEIEPLTT
FPMYPSLFPLRLKRPWHAGTSSLPDGRGDLGSLTWLRGGGGEQQGLPLNYPVGLFPW
MQQRDLDSQMGTDNNQQYQAMLAAGLQNIIGGDPLRQQFVQLQEPHHQYLQQSASHNSDL
MLQQQQQQQASRHLMAHTQIMSENLPQONMRQEVSNQPAQQQQQLQPPDNAYLNAFKM
QNGHLQQWQQQSEMPSPSFMKSDFTDSSNKFATTASPASGDGNLLNFSITGQSVLPEQLT
TEGWSPKASNTFSEPLSLPQAYPGKSLALEPGNPQNPSLFGVDPDSGLFLPSTVPRFASS
SGDAEASPMSLTDSGFQNSLYSCMQDTTHELLHGAGQINSSNQTKNFVKVYKSGSVGRSL
DISRFSSYHELREELGKMFAIEGLLEDPLRSGWQLVFVDKENDILLGDDPWESFVNNVW
YIKILSPEDVHQMDHGESSGGLFPQNPHTL*

>G1478 (1..354)

ATGTGTAGAGGGTTTGAGAAAGAAGAAGAGAGAAGAAGCGACAATGGAGGATGCCAAAGA
CTATGCACGGAGAGTCACAAAGCTCCGGTAAGCTGTGAGCTTTGCGGCGAGAACGCCACC
GTGTATTGTGAGGCAGACGCAGCTTTCCTTTGTAGGAAATGCGATCGATGGGTCCATTCT
GCTAATTTTCTAGCTCGGAGACATCTCCGGCGCGTGATCTGCACGACCTGTCCGAAGCTA
ACTCGTCGATGTCTTGTGCGGTGATAATTTAATGTTGTTTACCGGAGATAAGGATGATA
GCAAGGATTGAAGACATAGTAGTGATCACAAAATTCCTTTGTGTTTCTCTGA

>G1478 Amino Acid Sequence (domain in aa coordinates: 32-76)

MCRGFEKEEERRSDNGGCQRLCTESHKAPVSCLENGENATVYCEADAFLCRKCDRWVHS
ANFLARRHLRRVICTTCKRLTRCLVGDNFNVVLPEIRMIARIEHSSDHKIPFVFL*

>G1496 (116..1123)

AAACCCACCAAATAACTCAGAGCTTTTTTGCATTTTTTCCATTCTCTATTTTGTGTTTGT
ACTTTTGGTCTCACTTTAAAAGATCATAAGTTGAAAGATTCTGCAGAGAACAATATGTT
GGAAGGTCTGTCTCTCAAGAAAGCTTGCTCTTAACTCTATGGACATGTCTGTACTTGA
AAGGCTTAAATGGGTACAACAGCAACAACAGCAACTGCAACAAGTTGTGTCCCATAGCAG
TAATAATTACCTGAACCTTCTTCAGATACCTTCAGTTCCATGGAAGCAACAATGATGAGTT
GTTGGAGAGTAGTTTCAGCCAATTTCAAATGCTTGGATCTGGTTTTGGACCAAACTATAA
CATGGGTTTTGGTCTCCACATGAATCCATTTCAAGAACAAGTAGCTGCCATATGGAACC
TGTGGATACAATGGAGTTTTGTTGAAGACCGGTGAAGAAACCAGAGCCGTTGCCTTGAA
GAACAAGAGAAAACAGAGGTTAAGACAAGGGAAGAGCAAAAGACAGAGAAGAAGATCAA
AGTAGAGGCTGAGACAGAGTCAAGCATGAAAGGAAAATCAAACATGGGAAACACTGAAGC
ATCTTCAGACACTTCAAAGGAGACATCGAAAGGAGCTTCAGAGAATCAGAAATTAGATTA
TATCCACGTGAGAGCTCGTCGAGGCCAAGCCACTGACAGACAGCTTAGCAGAAAGGGC
GAGAAGAGAAAAAGATCAGCAAGAAAAATGAAATATCTGCAAGATATTGTGCC'TGGATGCAA
TAAGGTCACAGGAAAAGCTGGTATGCTTGATGAGATCATCAATTATGTTCAATGTCTCCA
AAGACAAGTCGAGTTCCTGTGATGAAACTTGCTGTCTTGAACCCGGAAGTAGAGCTTGC
CGTGGAAGATGTATCCGTAAACAGGCTTACTTTACAAATGTAGTTGCTTCAAAGCAATC
AATAATGGTTGATGTGCCATTGTTTCCGTTAGACCAGCAAGGATCTCTAGATTTGTCTGC
GATAAACC CGAACCAACGACATCTATCGAAGCTCCATCTGGAAGCTGGGAAACTCAATC
ACAGAGTCTCTACAACACATCTAGCCTCGGTTTTTATTACTAAGCAAGATTATTGAAAC
AACATGGTTGACATCAATCAATCATCAAAATCAGAAGCAAAATCTATTACATTTGCTCAT
CAAAAGTAGTAATTCGAAATTTGGTTAATGCATTATCCTTTGATCCTTGTCTTCTGATAT
TTAAACCAGAAAGAACTGGAGATAGCAATCCAATGATCTTGTACCA

>G1496 Amino Acid Sequence (domain in AA coordinates: 184-248)

MLEGLVSQESLSLNSMDMSVLERLKWVQQQQQLQVVS HSSNNSPELLQLQFHGSNND
ELLESSFQFQMLGSGFGPNYNMGFGPPHESISRTSSCHMEPVDMEVLLKTGEETRAVA

LKNKRKPEVKTREEQKTEKKIKVEAETESSMKGKSNMGNTASSDTSKETS KGASENQKL
DYIHVRARRGQATDRHSLAERARREKISKMKYLQDIVPGCNKVTGKAGMLDEIINYVQC
LQRQVEFLSMKLAVLNPELELAVEDSVKQAYFTNVVASKQSIMVDVPLPPLDQQGSLDL
SAINPNQTTTSIEAPSGSWETQSQSLYNTSSLGFHY*

>G1526 (1..3090)

ATGGGAACGAAAGTCTCAGACGATCTTGTTTCCACCGTCAGATCAGTCGTGGGTTCGGAT
TACTCAGATATGGATATAATCAGGGCTTTACACATGGCGAATCATGATCCAACGGCTGCT
ATCAATATAATCTTCGACACTCCAAGTTTCGCCAAACCTGATGTAGCCACTCCTACCCCG
AGCGGCTCTAATGGAGGGAAGCGAGTTGATAGTGGATTAAAGGGCTGTACTTTTGGTGAC
AGCGGAAGTGTGGAGCGAATCATCGCGTGGAGGAAGAAAATGAGAGTGTAAATGGTGGA
GGAGAAGAGAGTGTTCAGGGAATGAGTGGTGGTTTGTGGTTGTCTGAATTGGCTGGG
TTATCGACATGTAAAGGAAGGAAATTGAAGTCTGGTGATGAATTGGTGTTCACGTTTCCG
CATAGTAAAGGATTAAAGCTGAGACTACGCCCTGGGAAGCGCGTTTGTGGGCGGGGAAGG
CCAGCTTTGCGTGGTGCCTTCGTATATCGTTAGGTTCTCTACAAAGGATTGAGGAGAGATT
GGTAGAATAACCAACGAGTGGGCTCGGTGTCTTCTACCACTTGTGAGAGACAAGAAAATT
AGGATAGAAGGCAGTTGCAAGTCGGCGCCTGAAGCTTTGAGCATCATGGATACAATTCTT
CTGTCTGTAAGCGTGTACATTAATAGTTCCATGTTTCAAAAAGCATAGTCCGACTTCATTT
AAGACAGCTAGTAATACGGCAGAGGAATCAATGTTCCATCCTCTCCCAAATCTCTTTCGG
TTACTCGGTTTGATCCCTTTAAGAAGGCAGAGTTTACTCCAGAGGATTTTACTCTAAG
AAGCGACCTTTGAGTTCGAAGGATGGTTCCTGCTATTCCTACTTCGTTGCTTCAATTAAAC
AAGGTCAAGAATATGAATCAAGATGCAAAACGGAGATGAAAATGAGCAGTGTATCAGCGAT
GGTGATCTTGATAACATTGTTGGTGTGGGGACAGTTCTGGATTAAAGGAAATGGAACT
CCACATACACTTCTGTGTGAGCTTCGTCCATACCAAAAGCAGGCACCTTCATTGGATGACC
CAACTGGAGAAAGGAAATTGCACTGATGAGGCAGCAACAATGCTTCACCCGTGTTGGGAA
GCATACTGTTTAGCAGACAAGAGGGAAGTGGTTGCTACCTGAATTCTTTACTGGTGAT
GCTACAATACACTTCCCTAGCACACTTCAAATGGCAAGAGGAGGAATATTAGCAGACGCA
ATGGGTCTTGGAAGACTGTAATGACCATATCCCTTTGCTTGCCCATCTCTGGAAGCT
GCATCAACTGGGTTTCTATGCCCCAACTATGAAGGAGACAAAGTGATCAGCAGTTCTGTA
GATGATCTCACTAGTCCCCGGTGAAGGCAACCAAAATTTCTAGGCTTTGATAAGAGGCTT
CTTGAACAAAAAAGTGTACTTCAAAATGGTGGTAACCTGATTGTATGTCGGATGACACTT
TTAGGACAGTGGAAAGACAGAGATTGAAATGCATGCAAAAGCCTGGGTCTCTATCTGTCTAT
GTTCACTATGGGCAAAGCAGGCCGAAGGATGCAAACTTCTTTCCAGAGTGATGTGGTA
ATCACCACATATGGAGTTCTAACATCCGAATTCGCAAGAGAACTCAGCAGACCATGAA
GGAATTTATGCAGTTTCGATGGTTTAGGATTGTTCTTGACGAGGCACATACCATCAAAAAC
TCAAAAAGCCAAATTTCTTGGCTGCTGCAGCTCTGGTTGCTGATAGGCGTTGGTGTCTT
ACGGGTACTCCTATTGAGAACAATCTGGAGGATTTATACAGCCTTCTACGGTTTTTGAGG
ATTGAACCATGGGGAACCTGGGCATGGTGGAATAAACTTGTCAAAAGCCATTGGAAGAG
GGTGATGAGAGAGGGTTAAAGCTAGTGCAGTCTATCTTAAACCTATCATGCTTAGGAGA
ACAAAGTCTAGCACAGACCGAGAAGGAAGGCCGATTCTTGTTCTACCCCTGCTGATGCA
CGGGTCATTTACTGTGAACCTTCGGAGTCTGAGAGGGATTCTACGACGCGCTATTTAAA
AGATCCAAGTCAAATTTGATCAATTTGTTGAACAAGGCAAAGTTCTTCATAACTATGCT
TCGATCCTGGAACTGCTTTTGCGTCTTCGACAATGTTGTGATCACCCATTTTGTAGTAATG
AGTCGAGGGGATACAGCGGAATACTCTGATCTGAATAAGCTTTCTAAACGTTTCCTTAGT
GGAAAGTCTTCTGGCTTAGAAAGGGAAGGAAAAGATGTACCGTCAGAGGCTTTTGTTCAG
GAGGTGGTAGAGGAACGCGCAAAGGAGAGCAAGGAGAGTGTCCAATATGCCTTGAAGCA
CTTGAGGATGCTGTATTAAACGCCATGTGCTCATAGATTATGTCGTGAGTGTCTCTGGCA
AGTTGGAGAAATTTACTTCTGGGTTATGTCCTGTGTGTAGGAACACTGTAAGCAAAACA
GAACTCATCACAGCACCACCGAAAGTAGATTCCAGGTTGACGTGGAAGAAATGGGGTG
GAATCATCGAAAATCACTGCTCTTCTGGAAGAGCTTGAAGGTCTTCGTTCTTCAGGCTCT
AAGAGCATCTCTTTAGCCAGTGGACCGCTTTCCTCGATCTCCTCCAAATCCCTCTCT
CGGAATAACTTTTCATTTGTCCGTCTTGATGGCAGCCTAAGTCAGCAGCAACGAGAGAAG
GTCTTTAAAGAAATTTCCGAAGATGGCAGTATCCTGGTACTGTTGATGTCTCTAAAGCT
GGTGGCGTTGGGATAAATCTAACAGCTGCGTCCAATGCTTTTGTGATGGATCCATGGTGG
AAGCCAGCGTAGAGGAACAAGCTGTTATGCGTATTATCGTATAGGGCAAACTAAGGAA
GTCAAAATCAGAAGATTCATCGTTAAGGGAACGGTTGAAGAGAGAAATGGAGGCGGTTGAG
GCGAGGAAGCAGAGAATGATCTCTGGGCTTTAACCGATCAAGAAGTACGAAGTGCACGT
ATAGAGGAACCTCAAGATGTTATTTACCTGA

>G1526 Amino Acid Sequence (domain in AA coordinates: 493-620, 864-1006)

MGTKVSDDLVSTVRSVVGSDYSDMDIIRALHMANHDPTAAINIIFDTPSFAKPDVATPTP
SGSNGGKRVDSGLKGCTFGDSGSGVGANHRVEENESVNGGGEESVSGNEWWFVGCSELAG
LSTCKGRKLKSGDELVFTFPHSKGLKPETTPGKRGFGRGRPALRGASDIVRFSTKDSGEI
GRIPNEWARCLLPVLRDKKIRIEGSCSAPEALSIMDTILLSVSVYINSSMFQKHSATSF
KTASNTAEESMFHPLPNLFRLLGLIPFKKAEFTPEDFYSKRPLSSKDGSIAIPTSLQLN
KVKNMNQDANGDENEQCI SDGLDNIVGVGDSSGLKEMETPHTLLCELRPYQKQALHWM
QLEKGNCTDEAATMLHPCWEAYCLADKRELVVYLSFTGDATIHFPSTLQMARGGILADA
MGLGKTVMTISLHLSHWKAASGTGLCPNYEGDKVISSSVDDLTSPPVKATKFLGFDKRL
LEQKSVLQNGGNLIVCPMTLLGQWKTEIEMHAKPGSLSVYVHYGQSRPKDAKLLSQSDVV
ITTYGVLTSFESQENSADHEGIYAVRWFRIVLDEAHTIKNSKSQISLAAAALVADRRWCL
TGTPIQNNLEDLYSLRFLRIEPWGTAWWNKLVQKPFEEGDERGLKLVQSILKPIMLRR
TKSSTDREGRPIVLPPADARVIYCELSERDFYDALFKRSKVKFDQFVEQGKVLHNYA
SILELLRLRQCCDHPFLVMSRGDTAEYSDLNKLSKRFSGKSSGLEREGKDVPSEAFVQ
EVVEBLRKGEQGECPICLEALEDAVLTPCAHRLCRECLLASWRNSTSGLCPVCRNTVSKQ
ELITAPTESRFQVDVEKNWVSSKITALLEEGLRSGSGSKSILFSQWTAFLDLLQIPLS
RNNFSFVRLDGTLSQQQREKVLKEFSEDGSILVLLMSLKAGGVGINLTAASNAFVMDPWW
NPAVEEQAVMRIHRIGQTKEVKIRRFIVKGTVEERMEAVQARKQRMISGALTDQEVRSAR
IEELKMLFT*

>G1543 (1..828)

ATGATAAACTACTATTTACGTACATATGCACATACACATATAAACTATATGCTCTATAT
CATATGGATTACGCATGCGTGTGTATGTATAAATAAAGGCATCGTCACGCTTCAAGTT
TGTCCTCTTTTATATAAACTGAGAGTTTCTCTCAAACCTTTACCTTTCTCTTCGATC
CTAGCTCTTAAGAACCCTAATAATTCATTGATCAAAATAATGGCGATTTTGCCGAAAAC
TCTTCAAACCTTGATCTTACTATCTCCGTTCCAGGCTTCTCTCATCCCCTCTCTCCGAT
GAAGGAAGTGGCGGAGGAAGAGACCAGCTAAGGCTAGACATGAATCGGTTACCGTCGTCT
GAAGACGGAGACGATGAAGAATTCAGTCACGATGATGGCTCTGCTCCTCCGCGAAAGAAA
CTCCGTCTAACCAGAGAACAGTCACGTCTTCTTGAAGATAGTTTCAGACAGAATCATACC
CTTAATCCCAAAACAAAAGGAAGTACTTGCCAAGCATTGTATGCTACGGCCAAGACAAATT
GAAGTTTGGTTTCAAACCGTAGAGCAAGGAGCAAATTGAAGCAAACCGAGATGGAATGC
GAGTATCTCAAAGGTGGTTTGGTTCAATTAACGGAAGAAAACACAGGCTCCATAGAGAA
GTAGAAGAGCTTAGAGCCATAAAGGTTGGCCCAACAACGGTGAACCTCTGCCTCGAGCCTT
ACTATGTGTCTCGCTGCGAGCGAGTTACCCCTGCCGCGAGCCCTTCGAGGGCGGTGGTG
CCGTTCCGCTAAGAAAACGTTTCCGCCGCAAGAGCGTGATCGTTGA

>G1543 Amino Acid Sequence (domain in AA coordinates: 135-195)

MIKLLFTYICTYTYKLYALYHMDYACVCMYKYKIVTLQVCLFYIKLRVFLSNFTFSSSI
LALKNPNNSLIKIMAILPENSSNLDLTISVPGFSSSPLSDESGGGGRDQLRLDMNRLPSS
EDGDDEEFSHDDGSAPPRKKLRLTREQSRLLEDSEFRQNHNLNPKQKEVLAKHLMLRPRQI
EWFQNRARRSKLKQTEMECEYLKRWFGSLTEENHRLHREVEELRAIKVGPTTVNSASSL
TMCPRCERVTPAASPSRAVVPVPAKTFPPQERDR*

>G162 (101..619)

AGACATACAACACCAAATCTTCTTCTTCAACACATATTCACTTTACAGCAAAAAAAA
ACGAGAGGTTCTCTTATTTCGTACCGTTTAGCAAACAAATGGGTGCGAGAAAGATCAA
GATGGAGATGGTTCAAGAGCATGAACACACGACAGGTTACCTTTCAAACGGAGGACTGG
TTTGTTCAGAAGGCGAGCGAGTTAGCCACGCTCTGCAACGCTGAGTTGGGCATCGTTGT
CTTTTCAACAGGAGCAAGCCTTTCTCTACGGGAAACCGAATCTTGATTCTGTTGCAGA
GCGATTTCATGAGAGAAATATGATGATTTCAGACAGTGGCGATGAAGAAAAAGTGGTAATTA
CAGGCCTAACTGAAGAGGCTGAGTGAACGTCTCGATTGTCTCAACCAAGAGGTTGAAGC
TGAGAAGGAACGAGGCGAGAAGAGTCAGGAGAAGCTTGAATCTGCTGGGGATGAGAGATT
CAAGGAGTCCATTGAGACGCTTACCCTCGATGAACCAATGAATACAAAGATAGGCTTCA
GACAGTCCATGGTAGGATTGAAGGTCAAGTCAATCACTTGCAGGCTTCGTCTTGCCTCAT
GCTTCTCTCCAGAAAATAGCTAGACCGACTTGTTAGAGTTACATTCTATTTTTGTATCA
GCCTACAGAACTTACCAACACATGAAAGTTATTGCTGGTGTAGAATTTCTGTCTATCTAT
GGGGTGTGACTTTCTATTGACATCAAATGAAATGTACCTGGAAATTTGTCTGTATTAA
TCTCAAGTGTACTTGTCTAAACTTGATCAGCTTTTTCGCAAAAAAAA

>G162 Amino Acid Sequence (domain in AA coordinates: 2-57)

MGRRIKMEMVQDMNTRQVTFSKRRTGLFKKASELATLCNAELGIVVFSPGGKPFSGYKGP

NLDSVAERFMREYDDSDSGDEEKSGNYRPKLKRLSERLDLLNQVEAEKERGEKSQEKLE
SAGDERFKESIETLTLELNEYKDRLQTVHGRIEQVNHLOQASSCLMLLSRK*

>G1640 (168..1196)

TTCCGCCAGATCCTTCTCTATATAAGGAAGTTCATTTCATTGGAGAGGTTTCGCTGACA
AGCTGCTCTAGCTTATCTGGTACCGTCGACCTCTCACTCAAGGGTCCAAAAGTGTTCCT
CTTTTTCAGTTTCTCTTCTCTTTTGGACAGAAGAGACCGAGAAGCAATGGGAAGGGCTC
CGTGTGTGAGAAAATCGGGTTGAAGAGAGGGAGATGGACAGCCGAGGAAGATGAGATCC
TCACCAAGTATATTAGACCAATGGTGAAGGTTCTTGGCGATCTTGCCTAAGAAAGCTG
GATTGTTGAGATGTGGAAGAGCTGTAGACTAAGGTGGATAAACTACTTAAGAAGAGACT
TAAAAAGAGGAAATATTACTTCCGACGAAGAAGAAATAATCGTCAAGTTGCATTCCCTTC
TCGGCAACAGATGGTCACTTATTGCAACACATCTACCAGGAAGAACAGACAACGAAATTA
AAAACATTTGGAACCTACATCTCAGCCGCAAAATCTATGCCCTTCACTGCCGTTCCGGAG
ATGGACACAATCTACTCGTCAACGATGTAGTCTTGAAGAAATCTTGTTCATCGTCTTCTG
GAGCCAAGAAACAATAACAAGACCAAGAAGAAGAAGGAAGGACTAGTAGGTATCCA
TGAAGAAACACAAGCAATGGTGACGGCCTCACAATGTTTCTCACACCTAAGGAGCTAG
AGAGTGATTTCAGTGAGGGAGGGCAAAATGGTAATTTTGAAGGAGAGTCTTGGGGCCTT
ATGAGTGGTTGGATGGTGAGTTAGAACGGCTCTTGAGTAGTTGTGTCTGGGAATGCACTA
GTGAAGAGGCTGTGATTGGAGTAAATGATGAAAAGGTGTGTGAGAGTGGGGACAATAGTA
GTTGTTGTGTTAATTTGTTTGAAGAAGAACAAGGAAGCGAGACAAAGATTGGTCACGTAG
GAATCACAGAGGTTGATCATGATATGACGGTGGAAAGAGAAAGAGAGGGAAGTTTTTAA
GTTTCAATTCAAATGAAAATAATGATAAAGATTGGTGGGTGGTCTATGTAATTTCTCAG
AAGTTGGGTTTGGGGTTGATGAGGAGTTGCTTGATTGGGAGTTTCAAGGTAATGTCACTT
GTCAAAGTGATGATCTATGGGATCTCTCAGATATTGGAGAGATAACATTGGAGTGATTGT
ACCGAGCAAGTGGATTGGCGGCCGCTCTAGACAGGCCCTCGTACCGGATCTCTAGCTAGAG
CTTTCGTTTCGTATCATCGGTTTCGACAACGTTTCGTCAAGT

>G1640 Amino Acid Sequence (domain in AA coordinates: 14-115)

MGRAPCCEKIGLKRGRWTAEEDEILTKYIQTNEGSWRSLPKAGLLRCGKSCRLRWINY
LRRDLKRGNITSDEEIIIVKLHSLGNRWSLIATHLPGRTDNEIKNYWNSHLRKRKIYAFT
AVSGDGHNLVNDVVLKKSCTSSSSGAKNNNKTKKKKGRTSRSSMKKHQMVMTASQCFSQ
PKELSDSFSEGGQNGNFEGESLGPYEWLDGELERLLSSCVWECTSEEAVIGVNDEKVCES
GDNSSCCVNLFEEEQGSETKIGHVGITEVDHDMTVEREREGSFLSSNSNENNDKDWVGL
CNSSEVGFVDEELLDWEFQGNVTCQSDDLWDLSDIGEITLE*

>G1644 (1..348)

ATGAAATTGATTGATTGGAAGACTGTGCTTTGATGACTTACACCGAACTCATTTTGGGT
TTCTGCAATGTTTAAATGTTGATCTGCAGGAGGACTAGTGGACCTATGAGACGAGCAAAA
GGTGGTTGGACTCCAGAGGAGGATGAGACACTTAGACGAGCAGTTGAAAAGTATAAGGGG
AAGAGGTGGAAGAAAATAGCGGAATTTTCCCAGAGAGAAACACAAGTCCAATGCTTGCAC
AGGTGGCAGAAAGTTCTTAATCCAGAGCTTGTTAAAGGACCTTGGACTCAAGAGGTTCTC
TTATCATTTTTCATGTTCTGAAACTTTTTTTGGTTTTTCATTTTACGTAA

>G1644 Amino Acid Sequence (conserved domain in AA coordinates: 39-102)

MKLIDWKDCALMTYTELILGFCNVLM LICRRTSGPMRRAKGGWTPPEDETLRRAVEKYKG
KRWKKIAEFFPERTQVQCLHRWQKVLNPELVKGPWTQEVLLSFSCSETFFGFHFT*

>G1646 (34..786)

GATCTTTTGATCCAATCACAAGGCAAAGATCCAATGGACAATAACAACAACAACAAC
CAGCAACCACCACCAACCTCCGTCTATCCACCTGGCTCCGCCGTCAACAACGTAATCCCT
CCTCCACCATCTGGATCTGCATCAATAGTCACCGGAGGAGGAGCGACATACCACCACCTC
CTCCAGCAACAACAGCAAAGCTTCAAAATGTTCTGGACATACCAGAGACAAGAGATCGAA
CAGGTAAACGATTTCAAAAACCATCAGCTCCCTCTAGCTCGTATCAAAAAATCATGAAA
GCTGATGAAGATGTGCGTATGATCTCCGCCGAAGCACCGATTCTTTCGCGAAAGCTTGT
GAGCTTTTTCATTCTCGAACTTACGATTAGATCTTGGCTTCACGCTGAAGAGAACAAACGT
CGTACGCTTCAGAAAAACGATATCGCTGCTGCGATTACTAGAACCGATATCTTCGATTTC
CTTGTTGATATTGTTCTAGGGAAGAGATCAAGGAAGAGGAAGATGCAGCATCGGCTCTT
GGTGAGGAGGATGTTGCTCCCGCCGAGCGGTGTTCTTATTATTATCCACCGATG
GGACAACCGCGGTTCTTGGAGGGATGATGATTGGAAGACCGGCGATGGATCCTAGCGGT
GTTTATGCTCAGCCTCCTTCTCAGGCATGGCAAAGCGTTTGGCAGAATTACAGCTGGTGGT
GGTGATGATGTGCTTATGGAAGTGGAGGAAGTAGCGGCATGGTAATCTCGATAGCCAA
GGGTAAGTGAATTCTAGTAG

>G1646 Amino Acid Sequence (domain in AA coordinates: 72-162)
MDNNNNNNNQPPPTSVYPPGSAVTTVIPPSPSGSASIVTGGGATYHHLQOQQOQLQMF
WTYQRQEIEQVNDFNHQLPLARIKKIMKADEDVRMISAEAPILFAKACELFILELTIRS
WLHAEENKRRTLQKNDIAAAITRTDIFDFLVDIVPREEIKEEEDAASALGGGGMVAPAAS
GVPIYYPPMGQPAVPGGMMIGRPAMDPSGVYAQPFSQAWQSVWQNSAGGGDDVSYGSGGS
SGHGNLDSQG*

>G1672 (239..1399)

CCATTCTGACGTCCGGGATGACGCACAATCCCACTATCCTTCGCAAGACCCCTTCCTCTA
TATAAGGAAGTTCATTTTCATTTGGAGAGGACACGCTGACAAGCTGACTCTAGCAGATCTG
GTACCGATCACTCCCGTCTTTATCAAATCTTCTTCTCTTACATTTTCCCTATCCAATC
GATCTCACGCAGATCTGATCAATTTCTCATCAAATCATTTAGAGATCAAAAGAAAACCTAT
GAAGAATAGTAAATGTAACCTCATAGATTCAAAGCTCGAAGAACATCATCATCTTTGCGG
ATCAAAACATTGTCTGGATGTGGTCGCATGATTCAAGCTGCTACTAAACCAAATTGGGT
TGGATTGCCCGGCAGGATGAAATTCGATCCGACAGATCAAGAATTATAGAACATTTAGA
AGCAAAAGTGAAGGGAAGAAAGAAAATAAGAAATGGTTCGTCTCATCCACTTATAGA
TGAATTTATTTCCACCATTGATGGAGAAGATGGAATATGTTACACTCATCTCAGAAGCT
TCCAGGGGTGACAAGAGATGGCTTGAGCAAACACTTCTTCCACAAACCATCAAGAGCTTA
CACAAACCGGAACAAGAAAACGACGTAAATAATTCAAACCGATCACGACTCTGAGTTAAC
CGGATCATCAGAAACAGGTGGCAGAAAACGGGCAAAACAAGACCGGTTATGATCAACGG
TCAACAAAGAGGATGCAAGAAGATATTAGTACTCTACACAAACTTCGGCAAGAATCGTCG
ACCGGAGAAAACAAATTGGGTGATGCATCAATATCATTTAGGGATTAAATGAGGAAGAGAG
AGAAAGGAGAACTTGTGGTCTCCAAGATATTTTATCAGACACAACCAAGACAGTGTGTTAG
TAATACTAATTGGTCTGATCACCATGGTTCCAAGGACGTGATCGGAATTGGTGTGCGGAGA
TGAGATTTCCAGCGTAGCTGCCACGTTGCAGAGTCTTGGCTCCGGTGACGTCTTCTAG
GGTTAATATGCATCCCATACAGATCCTTTGATGAGGGGACAGCCGAAGCTTCAAAGGG
AAGAGAGAACCAGCATGTGTCTGGCACGTGCGAGGAAGTACATGATGGGATCATAACATC
ATCAATGTCTCATCATATGATTCATGATCATCATAATCAACATCATCAAATCGGAGA
TAGAAGAGAATTTACATGTCTCATCATATCCCATGACCCCTACTATCACATCAACA
TGAGTCAATCTTCCATGTTTACAAGTACTATGCCCTTTCAGCGGCAGCAATTAAGGGGTG
GTCTGTCTGGTTTCGGGATTAGAAGACCTAATTATGGGTTGTACCACAGCTACGTGTACAGA
AGACAATAATCACAATGATTAAATTCGCAGGAGCATTAGAAGCAAACCCCTCAGCGAAA
TGCAGAGTGGTTAACGTTTCCACAATTCTGGAACCAAGCCGAATCAGATGATCAAAACCG
AAGATTTTACAGAACCAAAAGGAAGCAGAGAAATCTTGCAAAAAGCTCCTGCTTAGCTG
TTGATCAATGCCGAAATGCTGAGCTATGACTGACTAGTCTCTGCCATTTAACTTACAAT
ATCACCAGAGGTTGCGATGAATGTTGATTGCTCAAAGGAGAGCGGCCGCTCTAGACAGG
CCTCGTACCG

>G1672 Amino Acid Sequence (conserved domain in AA coordinates: 41-194)

MKNSKCNLIDSKLEBHHHLCGSKHCPGCGRMIQAATKPNWVGLPAGVKFDPDQELIEHL
EAKVKGEENKKWSSSHPLIDEFIPTIDGEDGICYTHPQKLPGVTRDGLSKHFFHKPSRA
YTTGTRKRRKIIQTDHDSLTGSSETRWHKTGKTRPVMINGQQRGCKKILVLYTNFGKNR
RPEKTNWVMHQYHLGINEEBEREGLVVSKEIFYQTQPRQCVSNNTNWSDDHGSKDVGIGVG
DEISSVAATLQSLGSGDVVSRVNMHPHTRSFDEGTAEASKGRENQHVSGTCEEVHDGIIIT
SSMSSHHMIHDHNQHHQIGDRREFHSSSYPMPTTITSQHESIFHVTSTMPFQRQQLRG
RSSGSGLEDLIMGCTTATCTEDNNHK*

>G1677 (24..1037)

CAGTACTAATTTCTGTGTGTGTTAATGGTTCTAGTTATGGATGATGAAGAGAGTAACAACG
TTGAAAGATATGACGACGTGCTATTGCCAGGGTTTAGGTTCCATCCCACTGATGAAGAAC
TCGTAAGTTTCTACTTGAAACGGAAGGTTTACACAAATCTTCTCCCTTTGATCTCATCA
AGAAAGTCGACATTTACAAATACGATCCATGGGACCTCCCAAAGCTTGACGCGATGGGGG
AAAAAGAGTGGTACTTTTATTGTCTAGAGACAGGAAATACCGCAACAGCACAAAGACCTA
ACCGAGTAACTGGAGGTGGCTTCTGGAAAGCAACCGGAACAGACCGGCCTATATACTCAT
TGGACTCCACTCGATGCATCGGTTTGAAGAAATCACTTGTGTCTACCGTGGTCGAGCTG
CTAAAGGATCAAAACCGATTGGATGATGCATGAATTCGTCTCCCTTCTCTCTGACT
CTCATCACTCATATATCCCAATTACAATAACAAGAAGCAACACCTTAACAATAACAACA
ACAGCAAGGAGCTTCTTCAAACGATGCTTGGGCGATATGTAGAATATTTAAGAAGACAA
ATGCAGTATCCTCACAAGATCAATCCACAATCTTGGGTTTATCCAACGATTCCTGACA
ACAATCAACAGTCAACAACAACACCGCAACTCTCTTAGCTTCATCAGACGTTCTCAGCC

ACATATCAACAAGACAAAACCTTTATTCCTTCTCCAGTCAACGAACCCGCAAGCTTCACAG
AATCAGCTGCTTCTTACTTCGCGTCTCAGATGCTCGGAGTCACGTACAATACAGCCAGAA
ACAACGGAACAGGGGATGCTCTGTTTCTGAGAAACAATGGAACAGGGGATGCTCTGGTTC
TGAGCAACAATGAGAATACTACTTCAACAACCTTGACTGGAGGGTTGACTCATGAGGTTT
CGAATGTAAGATCAATGGTGATGGAGGAGACTACGGGGAGTGAGATGTCGGCGACGTCGT
ATTCCACTAACAATTAAGATCATAGTACTATTAACACTTGAATTAGTGTAGACGTTGATC
ATCGCTAATATGTATTAATTTTTCTTGCTTACTATAAACGAAAAA

>G1677 Amino Acid Sequence (conserved domain in AA coordinates:17-181)

MVLVMDDEESNNVERYDDVLPGRFPHPTDEELVSFYLKRKVLHKSPLFDLIKVDIYKY
DPWDLPKLAAMGEKEWYFYCPRDRKYRNSRPNRVTTGGFWKATGTDRIYSLDSTRCIG
LKKS LVFYRGRAAKGVKTDWMMHEFRLPSLSDSHSSYPNYYNKKQHLNNNNNSKELPSN
DAWAICRIFKKTNAVSSQRSIPQSWVYPTIPDNNQOSHNNATLLASSDVLSHISTRQNF
IPSPVNEPASFTESAASYFASQMLGVTYNTARNNGTGDALFLRNNGTGDALVLSNNENNY
FNNLTGGLTHEVPNVRSMVMEETTGSSEMSATSYSTNN*

>G1765 (139..966)

TCCTTCGCAAGACCTTCTCTATATAAGGAAGTTCATTTCAATTTGGAGAGGACACGCTG
ACAAGCTGACTCTAGCAGATCTGGTACCGTCGACAAGAATGACTTGATTGGTGTTCTAAA
GAGATCGATGTAGTGAAGATGAGTGGCGAAGGTAACCTTAGGTAAGGATCATGAAGAAGAA
AACGAAGCACCCTTCTGGGTTTCAGGTTTCATCCGACGGATGAAGAGCTTTTAGGATAC
TATCTTCGAAGAAAAGTAGAGAACAAAACCATCAAACTCGAAGCTTATCAACAGATCGAT
ATCTATAAGTACGATCCTTGGGATCTTCCAAGAGTGAGCAGCGTCGGAGAAAAGGAGTGG
TACTTCTTCTGCATGAGAGGTAGGAAATACAGGAATAGCGTTTCGACCAACCGAGTGACC
GGTTTCAGGTTTCTGGAAGCCACTGGTATTGATAAACCGGTTTACTCCAATCTTGACTGT
GTTGGTCTCAAGAAATCTCTGGTTTACTATCTTGGTTCAGCCGGTAAAGGCACCAAAACC
GATTGGATGATGCATGAATTCGCGCTCCCTCCACCACGAAAACCGACTCTCCAGCTCAA
CAAGCAGAGGTATGGACACTTTGCAGAATCTTCAAACGAGTCACATCTCAAAGAAACCCA
ACCATCTTACCACCAAAACCGAAAACCGGTTATCACTTTAACCGACACTTGTCTAAGACC
AGCAGCTTAGATTCCGACCACACGAGCCACCGTACAGTAGATTCCATGTCCACGAGCCG
CCGCTTCCACAGCCACAGAACTTATTGGAACCAACATATAGTTGGTTTTAATCAACCG
ACATATACTGTGTAATGATAATAACCTCTCTGATGAGTTTCTGGAACGGCAACGGTGGAGAT
TTCATAGGAGACTCAGCAAGTTGGGATGAACTTAGATCTGTTATAGATGGCAACACTAAA
CCCTAGTAATAAAAGTTTCTTTTTTTCAGCTTTGTACAAAAAGATAAAACAAACGGCAACC
GCTCTAGACAGGCTCGTACCGGATCTCTAGCTAGAGCTTTCGTTTCGTATCATCGGT
TTCGACAACGTTCTGT

>G1765 Amino Acid Sequence (conserved domain in AA coordinates: 20-140)

MSGEGNLGKDHEEENEAPLPGRFPHPTDEELLGYLLRRKVENKTIKLELIKQIDYKYDP
WDLPRVSSVGEKEWYFFCMRGRKYRNSVRPNRVTTGSGFWKATGIDKPVYSNLDVCVGLKKS
LVYYLGSAGKGTKTWMMHEFRLPSTTKTDSAPAQAEVWTLCRIFKRVTSQRNPITLPPN
RKPVITLTDTCSTSSLDSDHTSHRTVDSMSHEPPLPQPQNPYWNQHVGNQPTYTGND
NNLLMSFWNGNGGDFIGDSASWDELRSVIDGNTKP*

>G1777 (97..1878)

CTCGTACTTTATCACCTCCGTCGTTCTATAATACTCTCTTCCGTCATCATATCATTGT
CGACAATTTCAATCTGATCAGTTTAAAAATTGATCCATGGATGATAATTTAAGCGCGAG
GAAGAAGATTACTATTACTCCTCCGATCAGGAATCTCTCAACGGGATTGATAATGATGAA
TCCGTTTCGATACCTGTTTCTTCCGATCAAATACTGTCAAGGTTATTACGAAGGAATCA
CTTTTGGCTGCACAGAGGGAGGATTTCGCGAGAGTGATGGAATTGTTATCGGTTAAGGAG
CACCATGCTCGGACTCTTCTTATACATTACCGATGGGATGTGGAGAAGTTGTTTGCTGTT
CTTGTTGAGAAAGGSAAGATAGCTTGTCTTCTGGTGCTGGTGTACACTTCTTGAAAC
CAAAGTTGTGATTCTTCCGTTTCTGGTCTTCTTCGATGATGAGTTGTGATATCTCGGTA
GAGGATGTACCGGTTTATCAGCTGACAAGGATGGACTGTGGCCATAGCTTTTGCAATAAC
TGTTGGACTGGGCATTTTACTGTAAAGATAAATGAAGGTCAGAGCAAAAGGATTATATGC
ATGGCTCATAAGTGTAATGCTATTTGTGATGAAGATGTTGTCAGGGCTCTAGTTAGTAAA
AGCCAACAGATTTAGCTGAGAAGTTTGATCGTTTTCTTCTTGTAGTCATATCGAAGAT
AACAAAATGGTGAGGTGTCGAGTACTCCTCATTGTGGGAATGCCATACGTGTTGAG
GATGACGAGCTCTGTGAGGTTGAATGCTCTTGTGGTTTGCAGTTCTGTTTCAGTTGTTCA
TCTCAAGCTCACTCCCTTGTCTTGTGTGATGTGGGAACATATGGAGAAAGAAGTGCTTT
GATGAGTCCGAGACTGTTAATTGGATAACTGTTACACAAAGCCGTGTCCCAATGTAC

AAGCCTGTTGAAAAGAATGGTGGATGCAATCTCGTGACTTGTCTTTGTCGACAATCTTTT
TGTTGGTGTGTGGTGAAGCTACTGGAAGGGACCACACTTGGGCTAGAATCTCGGGTCAT
AGTTGTGGTTCGGTTCCAAGAAGATAAAGAGAAACAAATGGAGAGAGCGAAAAGGGATCTC
AAGCGGTATATGCATTATCATAACCGATACAAAGCACATATCGACTCCTCCAAGCTAGAG
GCTAAGCTTAGTAATAATATTAGTAAAAAGGTGTCTATTTAGAAAAGAGGGAGTTACAA
CTTAAAGACTTCAGCTGGGCTACCAATGGACTCCATCGGTTATTTAGATCAAGACGAGTT
CTTTCATATTATACCTTTTCGCATTTTACATGTTTGGAGATGAGCTGTTTAAAGATGAG
ATGAGCTCTGAGGAAAGAGAAATAAAACAAAATCTGTTTGAGGATCAGCAGCAGCAGCTT
GAGGCTAATGTTGAGAACTTTCTAAGTTCTTGGAGGAACCTTTTGATCAATTTGCTGAT
GATAAGGTCATGCAGATAAGGATTCAAGTCATCAATTTGTGAGTTGCGGTCGATACACTC
TGCGAAAAATATGTATGAATGCATTGAGAATGACTTGTGGGTTCTCTGCAACTTGGCATC
CACAACATTACTCCATACAGATCAAACGGCATAGAACGAGCATCTGATTTTATAGTTCC
CAGAAATCCAAGGAAGCTGTGGTCAGTCTTCGGATTGTGGATGGACGTCCAGGCTCGAT
CAAGCTTTGGAGTCAGGGAAGTCGGAAGACACAAGTTGCTCTTCCGGGAAGCGTGCTAGA
ATAGACGAAAGTTACAGAAACAGCCAAACCACCTTACTAGATTTAACTTGCCAGCGGAA
GCCATTGAGCGGAAATGAACACTTATCCTTCTTACCTCCCAATAACACCTTTTGTCC
AAATAAAGTGTGTTACCCGGATATTTATAGCTCTAAACCAATCCCCTCTGCTTAATTTG
TCAGTGACCTTACCTAACCTCTTCA
>G1777 Amino Acid Sequence (domain in AA coordinates:124-247)
MDDNLSGEEEDYYYSSDQESLNGIDNDESVSIPVSSRSNTVKVITKESLLAAQREDLRRV
MELLSVKEHHARTLLIHYRWVDEKLFVAVLVEKGKDSLFSGAGVTLLLENQSCDSSVSGSSS
MMSCDICVEDVPGYQLTRMDCGHSFCNNCWTHGFTVKINEGQSKRIICMAHKCNAICDED
VVRLVSKSQPDLAEKFDRLLESYIEDNKMVKWCPSTPHCGNAIRVEDDELCEVECSG
LQFCFSCSSQAHSPSCVMWELWRKKCFDESETVNWITVHTKPCPKCHKPVEKNGGCNLV
TCLCRQSFCLWLCGEATGRDHTWARI SGHSCGRFQEDKEKQMERAKRDLKRYMHYHNRYKA
HIDSSKLEAKLSNNISKVSISEKRELQLKDFSWATNGLHRLFRSRRVLSYSYPFAFYMF
GDELFKDEMSSSEEREIKQNLFFEDQQQLEANVEKLSKFLPEPFDQFADDKVMQIRIQVIN
LSVAVDTLCEMYECIENDLLGSLQLGIHNITPYRSNGIERASDFYSSQNSKEAVGQSSD
CGWTSRLDQALESGKSEDTSCSSGKRARIDESYRNSQTLLDLNLP AEAI ERK*
>G1793 (59..1783)
AGTGATTTATTGATTAACCCAAACACAAAATAAACAGATTTGACTCAAAAAGAAGAAAAT
GAATTCCTAACAACTGGCTTGGCTTTCCTCTTTCACCGAACAACTCTTCTTGCCTCCTCA
TGAATACAACTTGGCTTGGTCAGCGACCATATGGACAACCTTTTCAAACACAAGAGTG
GAATATGATCAATCCACACGGTGGAGGAGGAGATGAAGGAGGAGAGGTTCCAAAAGTGGC
CGATTTTCTCGGTGTGAGCAAACCGGACGAAAACCAATCCAACCACCTAGTAGCTTACAA
CGACTCAGACTACTACTTCCATACCAATAGCTTGATGCCTAGCGTCCAATCAAACGATGT
CGTTGTAGCAGCTTGTGACTCCTAATACTCCTAACAACAGTAGCTATCATGAGCTTCAAGA
GAGTGCTCACAATCTACAGTCACCTTACTTTGTCCATGGGGACCACCGCTGGTAATAATGT
TG TAGACAAAAGCTTCAACATCCGAGACCACCGGGGATAACGCTAGCGGTGGAGCACTAGC
CGTTGTGAGACGGCCACGCCAAGACGTGCATTGGACACTTTCGGACAACGAACCTCGAT
CTATCGTGGTGTCAACAAGACATCGATGGACTGGTCGATATGAGGCTCATCTATGGGATAA
TAGTTGTAGAAGGGAAGGCCAGTCTAGGAAAGGAAGACAAGTTTACTTGGGTGGATATGA
CAAAGAAGATAAAGCAGCAAGATCATATGATCTAGCTGCCTTAAGTACTGGGGTCTTTC
AACTACTACTAATTTCCCATTAACAACTACGAGAAAGAGTAGAGGAAATGAAGCATAT
GACGAGACAAGAGTTCTGTGGCTGCCATTAGAAGGAAAAGTAGTGGATTTTCGAGAGGCGC
TTCGATGTATCGAGGAGTTACAAGGCATCACCAACATGGAAGATGGCAAGCAAGGATCGG
CCGAGTCGCGCGAAACAAAGACCTCTACTTGGGAACTTTTCAGCACTGAGGAAGAAGCAGC
AGAAGCTTACGATATAGCTGCAATAAAGTTTAGAGGACTTAATGCAGTGACCAACTTCGA
GATCAACCGGTACGACGTGAAAGCCATTCTAGAGAGTAGCACTCTTCCCATCGAGGAGG
CGCAGCTAAACGGCTCAAAGAAGCTCAAGCTCTTGAGTCTTCAAGGAAACGCGAGGCGGA
GATGATAGCCCTTGGTTCAAGTTTCCAGTACGGTGGTGGCTCGAGCACAGGCTCTGGCTC
CACCTCATCAAGACTTCAGCTTCAACCTTACCCTCTAAGCATTCAACAACCATTAGAGCC
TTTCTATCTCTTCAGAACAAATGACATCTCTATTACAACAACAACATGCTCAGGATTC
CTCCTCTTTTATACCATAGCTATATCCAGACACAACCTTCATCTCCACCAACAGACCAA
CAATTACTTGCAGCAACAGTCGAGCCAGAACTCTCAGCAGCTCTACAATGCGTATCTTCA
TAGCAATCCGGCTCTGCTTCATGGACTTGTCTCTACCTCTATCGTTGACAACAATAATAA
CAATGGAGGCTCTAGTGGGAGCTACAACACTGCAGCATTCTTGGGAACCAAGGATTTGG

TATTGGGTCCAGCTCGACTGTTGGATCGACCGAGGAGTTTCCAACCGTTAAAAACAGATTA
CGATATGCCTTCCAGTGATGGAACCGGAGGGTATAGTGGTTGGACCAAGTGAGTCTGTTCA
GGGGTCAAACCTGGTGGTGTTTTCACTATGTGGAATGAGTAAACAAGGATCTCTTTCTT
GCGGCACAAGGAATGGGT

>G1793 Amino Acid Sequence (conserved domain in AA coordinates:179-255, 281-349)

MNSNNWLGFPLSPNNSSLPPHEYNLGLVSDHMDNPFQTQEWNNMINPHGGGGDEGGEVVKV
ADFLGVSKPDENQSNHLVAYNDSYYFHTNSLMPSVQSNVVAACDSNTPNNSSYHELO
ESAHLNQLSLTSMGTTAGNNVVDKASPSETTGDNASGGALAVVETATPRRALDTFGQRTS
IYRGVTRHRWTRGRYEAHLWDNSCRREGQSRKGRQVYLGYYDKEDKAARSYDLAALKYWGP
STTTNFPITNYEKEVEEMKHMTRQEFVAAIRRKSSGFSRGASMYRGVTRHHQHGRWQARI
GRVAGNKDLYLGTFFSTEBEAAEYDIAAIKFRGLNAVTFEINRYDVKAILESSTLPIGG
GAAKRLKEAQALESSRKREAEMIALGSSSFQYGGSSSTSGSSTSSRLQLQPYPLSIQPLE
PFLSLQNNDISHYNNNNNAHDSSSFNHHSYIQTQLHLHQQTNNYLQQSSQNSQQLYNAYL
HSNPALLHGLVSTSIVDNNNNNGSSGSYNTAAFLGNHIGIGSSSTVGSTEEFPTVKTD
YDMPSSDGTGGYSGWTSSESVQGSNPGGVFTMWNE*

>G180 (54..629)

GTAATTACGATCTACAACAAGTGACATCGTCGTCGACGACGATTCAAGAGAATATGAACT
TCCTCGTTCCCTTTTGAAGAAACCAATGTCTTAACCTTTTCTCTTCTTCTTCTTCTTCTT
CTCTTTCTTCTCTTCTTCTTCCCCATTCACAACTCTTCTCCACTACTACTACTCATGCAC
CTCTAGGGTTTCTAATAATCTTCAGGGTGGAGGACCTTGGGATCAAAGGTGGTTAATG
ATGATCAGGAGAATTTGGAGGTGGAATAACAATGATGCTCATTCTAATCTTGGTGGA
GATCAAATAGTGGAGGTGGAGATATGAAGAACAAGTGAAGATAAGGAGGAACTAAGAG
AGCCAAGATTCTGTTTCCAAACCAAAAGCGATGTTGATGTTCTTGACGATGGCTACAAAT
GGCGTAAATATGGTCAGAAAGTCGTCAAGAACAGCCTTCACCCCAGGAGTTATTACAGAT
GCACACACAACACTGTAGGGTGAAAAAGAGAGTGGAGCGACTATCGGAAGATTGTAGAA
TGGTGATTACTACTACGAAGGTCTGTCAACACCATTTCCCTCTGATGACTCCACTTCTC
CTGACCATGATTGTCTCTTCTTCTTTTAAACATCTCTTCTATATATCTATATATAGACAG
TTATATGTGCACATATAGATGTGTGATATATTGCATATTTGATATTGCATGTGTTTTC
AGAGTATGTCATCAGATGTTATGCATATATTCTTGACTTGTGCTTATAGTATACATATG
TAATAATATATATTGACATTGGTAGTTTCTGTTTCTGTTCAAACAAAAA

>G180 Amino Acid Sequence (domain in AA coordinates: 118-174)

MNFLVPFEETNVLTFPSSSSSSSLSSPSFPIHNSSTTTTHAPLGFSNNLQGGGPLGSKV
VNDDQENFGGGTNNDAHSNSWRSNSGSGDMKNVKIRKLRPRFCFQTKSDVDVLDG
YKWRKYGQKVVKNSLHPRSYYRCTHNNCRVKRVERLSEDCRMVITTYEGRHNHIPSDDS
TSPDHDCLSSF*

>G192 (63..959)

CTTTTTTCTCTTCTCTCTCCTCAGAGATTGGAAGCTTTTTGTCTCCCTGAGTAACCAAATT
CAATGGCCGAGATTGGGATCTCCACGCCGTAGTCAGAGGCTGCTCAGCCGTAAGCTCAT
CAGCTACTACCACCGTATATTCCCCGGCGTTTCATCTCACACAAACCTATATTCACCG
TCGGACGACAAAGTAATGCCGTCTCTTCGGAGAGATTGAGATCTCTACACACCGTTCA
CACAAGAATCTGTCGTCTCTTCGTTTTCTGTATAAACTACCCAGAAGAACCTAGAAAGC
CACAGAACCAGAAACGTCTCTTCTCTCTCTGCTTCTTCGGTAGCGTCACTAGCAAAC
CCAGTGGCTCCAATACCTCTAGATCTAAAAGAAGAAAGATACAGCATAAGAAAGTGTGCC
ATGTAGCAGCAGAAGCTTTAAACTCCGATGTCTGGGCATGGCGAAAGTACGGACAGAAAC
CCATCAAAGGTTACCATATCCAAGAGGATACTACAGATGTAGTACATCAAAGGTTGTT
TAGCCCGTAACAAGTGGAGCGAAATAGATCCGACCCGAAGATGTTTATCGTCACTTACA
CGGCGGAGCATAATCATCCAGCTCCGACACACCGTAATCTCTCGCCGGAAGCACACGTC
AGAAACCATCCGATCAACAGACGAGTAAATCTCCGACGACCACTATTGCTACTTATTCAT
CGTCTCCGGTGACTTCAGCCGACGAATTTGTTTGCCTGTTGAGGATCATCTAGCGGTGG
GAGATCTTGACGGAGAAGAGATCTGTTATCTTTGTCGGATACGGTGGTTAGCGATGATT
TCTTCGATGGGTTAGAGGAATTCGACGCCGAGATAGCTTTTCCGGGAACCTCGGCTCCGG
CGAGTTTGTATCTCTTGGGTTGTGAACAGTGCCGCCACTACCACCGGAGGAATATGAT
TAGATTACGACGGCTTAGAATACTCTTATTAGGACAGATTTATAGGATTAAGGAATTATT
CTCGGAGCATATGTAAAAATAGGATAAAAGAAAATGTTCTTTGTTACTTTTTTTCGGGTT
TTCTTCTTATGTTTCTAAACATCTTAGAAAAAATTAATTGTATATTCTTAAAGCTCGA
TACATCTGTTTTAAAAA

>G192 Amino Acid Sequence (domain in AA coordinates: 128-185)

MADDWDLHAYVRGCSAVSSSATTTVYSPVGSSTHTNPIFTVGRQSNVAFGEIRIDLYTPTFT
QESVVSFSCINYPEEPRKPQNQKRPLSLSSASSGSVTSKPSGSNTSRSKRKKIQHKKVCH
VAAEALNSDVWAWRKYGQKPIKGSYPYRGYYRCSTSKGCLARKQVERNRSDPKMFIVTYT
AEHNHPAPTHRNSLAGSTRQKPSDQQTSSKSPTTTTIATYSSSPVTSADFEVFLPVEDHLAVG
DLDGEEDDLSSLSTTVVSDDFDGLLEFAAGDSFSGNSAPASFDLSWVNVNSAATTTGGI*
>G1948 (18..1118)
AAAAGGTCTTCTTGCCATGGATACTTGTGCTCTAGTAATCCATCAGTCTCTGTCTCGCA
TCAAACCTTCTCTCCCAATCTTCTTCTTCTTCTTCTGCTTTCTCCCTGAATCCT
TACCACTCAGACGGAATCGAGCTGTGTTTCCGAGGAGCTATATGTGCCGCCGTACAAAGAA
ACTACGAAGAAACGACCTCCTCCGTGGAAGAGGCAGAGGAAGATGATGAGTCATCATCAT
CGTACGGAGAAGTGAACAAGATCATTGGAAGCCGAACGGCGGGGAAGGAGCCATGGAGT
ACCTTATCGAGTGAAGGACGGCCATTCTCCGTCTGTTGGTTCATCGAGCTACATCGCAG
CAGACGTAGTGTCTGAGTACGAGACACCTGGTGGACGCAGCTAGAAAAGCCGACGAGC
AGGCCTGTACAGCTCCTGGAGGACCGAGACGTCGATGCCGTGGACGAAAACGGCCGGA
CGGCTCTGCTTTTCGTGGCAGGTCTGGGGTTCGACAAAGTTCGCAAGGCTTCTGGCGGAGG
CTGGAGCCGATCTCGACACCGAGACATGAGGGGAGGCTTGACGGCGCTGCACATGGCGG
CTGGTTAGCTGAGGCCGAGGTGGTGGAGGCGCTGGTGGAGCTGGGAGCTGATATTGAAG
TGAAGACGAGAGAGGGTTAACGGCGTTGGAAGTAGCGAGGGAGATTCTGAAGACGACGC
CGAAGGGGAATCCGATGCGATTCTGGGAGGAGAATTGGGTAGAGAAAGTGATCAATGTCC
TGAAGGACAAGTGTTCTGAGTACGCCGAGGTGGATGAGATCGTAGAGAAACGAGGGAAAG
GCAAAGACGTTGAATATCTGGTCAGATGGAAGGACGGTGGAGATTGCGAGTGGGTGAAAG
GTGTACACGTGGCGGAAGATGTGGCTAAGGACTACGAGGATGGGCTGGAGTACGCTGTAG
CGGAGAGTGTGATCGGGAAGAGGGTGGGAGACGATGGGAAGACCATCGAGTATCTTGTC
AATGAGCTGATATGTCTGATGCCATTGGGAGCCTCAGGACAATGTCTGACTCTACTCTTG
TTCTACTCTACCAACAACAACAACCAATGAATGAATGATTGATTTTGATGATTACATTCT
TCTCAATTTGCTTCTTTCTCATATGTGTTGGTTCATCTGACCGGTTTCGGTTGGTACGTAC
CGGTACATTTTCATTTTCTTTAAGATGTGATCTTGATGTTTTCGGCTTTTGGGGACA
CTATTTGATTTTATATCCATGCTTTGAATTTGCTTCCCTTTTGGGGAGATTTCATGAAA
>G1948 Amino Acid Sequence (domain in AA coordinates: entire protein)
MDTCALVIHQSLSRILSPKSSSSSSSAFSPESLPPIRRIELCFRGAICAQVQRNYEET
SSVEEAEDDESSSSSYGEVNKLIIGSRITAGEGAMEYLIEWKDGHSPSSVPSYIADVVSE
YETPWWTAARKADEQALSQLEDDRVDVAVDENGRTALLFVAGLGSDKCVRLLEAGADLD
HRDMPWGLTALHMAAGYVRPEVVEALVELGADIEVEDERGLTALELAREILKTTPKGNPM
QFGRRIGLEKVINVLEGQVFEYAEVDEIVEKRGKGKDVEYLVRWKDGGDCEWVKGVHVAE
DVAKDYEDGLEAYAVAESVIGKRVGDDGKTIEYLVKWTDMSDATWEPQDNVDSTLVLLYQQ
QQPMNE*
>G2123 (1..657)
ATGAGAAAAGTATGTGAGCTTGATATAGAGCTAAGTGAAGAGGAAAGAGACCTACTAACA
ACTGGATACAAGAATGTCTGAGGCTAAGAGAGTTCATTGAGAGTAATATCATCCATT
GAAAAAATGGAAGACTCGAAAGGAAACGACCAAATGTGAAACTGATAAAAGGACAACAA
GAAATGGTTAAAGATAGTGTCTTCAATGTTTGAATGACATTTTGTCTCTCATTGATTCT
CATCTCATACCATCACTACTACTAATGTCTGAATCAATTGCTCTTTTAAACAGAGTGAAA
GGAGATTATTTTCGATATATGGCAGAGTTTGGTCTGATGCTGAACGTAAAGAAAATGCA
GATAATTCTCTAGATGCATATAAGGTTGCAATGGAAATGGCAGAGAATAGTTTAGACCCC
ACCAATATGGTTAGACTTGGATTGGCTTTAAATTTCTCGATATTCAATTATGAGATCCAT
AAATCTATTGAAAGCGCATGTAAATTGGTTAAGAAAGCTTACGATGAAGCAATCACTGAA
CTCGATGGCCTTGACAAGAATATATGCGAAGAGAGCATGTATATCATAGAGATGCTTTAA
TACAATCTTCTACCTGGACTTCAGGCGATGGTAATGGTAATAAGACAGACGGTTAG
>G2123 Amino Acid Sequence (domain in AA coordinates: 99-109)
MRKVCELDIELSEERDLLTTGYKNVMEAKRVSLRVISSIEKMEDSKGNDQNVKLIKQQ
EMVKYEFNVNCDILSLIDSHLIPSTTTNVEIVLFRNVKGDYFRYMAEFGSDAERKENA
DNSLDAYKVAMEAENSLAPTNMVRLLGLALNFSIFNYEIHKSIESACKLVKKAYDEATTE
LDGLDKNICESMYIIEMLKYNLSTWTSBGDNGNKTDG*
>G2138 (27..512)
GGAACCCATAATTTCCGCAAAATCACTATGAAGCGTATTATCAGAATCTCATTACCGACG
CAGAAGCCACCGATTCTTCTAGCGACGAAGACACGGAGGAGCGTGGAGGAGCATCCCAGA
CTCGGCGCGCTGGGAAACGCCCTCGTTAAAGAGATCGTAATCGATCCTTCCGATTCCGCCG

ATAAACTCGATGTCTGCAAAACACGGTTCAAAATCAGGATCCCGGCGGAATTTCTCAAGA
CGCGGAAAACGGAGAAGAAATATCGTGGAGTGAGGCAGAGGCCGTGGGGGAAGTGGGTGG
CGGAGATCAGATGTGGAAGAGGAGCTTGTAAGGACGACGTGATCGTCTCTGGCTGGGTA
CTTTTAACACTGCTGAGGAAGCTGCTCTAGCTTATGATAACGCTTCAATTAAGCTGATTG
GACCTCACGCGCCGACCAATTTTGGTTTGCCGGCGGAGAATCAAGAGGATAAGACGGTGA
TTGGAGCTTCTGAGGTTGCTAGAGGCGCGTGAAGTGGGGTTGGTAATTTAGTTGTTAGC
>G2138 Amino Acid Sequence (domain in AA coordinates: TBD)
MKRIIRISFTDAEATDSSDEDETERGGASQTRRRGKRLVKEIVIDPSDSADKLDVCKTR
FKIRIPAEFLKTAKTEKKYRGVRQRPWGKWWAEIRCGRGACKGRRDRLWLGTFTNTABEEA
LAYDNASIKLIGHPAPTNFGLPAENQEDKTVIGASEVARGA*
>G2139 (40..663)
CCTACAAGAAATCAAACACTAGTTCTGGTTTCTGCAAAACATGTCATCTACGAAGCAAGCA
AAGGGAAGAAAAACAAAGGGGAAGCAAAAGATCGAGATGAAGAAGGTGGAGAAGTATGGA
GATAGGATGATTACGTTCTCAAAACGTAACCGGAATTTTAAAGAAATGAACGAGCTC
GTAGCAATGTGTGACGTTGAAGTGGCTTTCTTGATTTTCTCTCAACCAAGAAGCCCTAT
ACATTCGCACATCCGCTCTATGAAGAAAGTGGCTGACCGGTTAAAGAACCCCTTCGAGACAA
GAACCATTAGAGAGAGACGATACCAGACCCCTCGTCGAAGCTTATAAGAAACGAAGGCTC
CACGACCTCGTAAAAAAATGGAGGCGCTCGAAGAGGAGCTTGCGATGGATCTAGAGAAG
TTGAAACTGTTGAAGGAATCGAGAAATGAAAAGAAGTTAGATAAAATGTGGTGGAACTTT
CCTTCGGAAGGTTTGAGCGGAAGGAGCTGCAGCAAAGGTACCAAGCGATGCTCGAGTTA
CGTGATAACTTATGCGACAATATGGCTCACTTACGATTGGGAAAAGACTGTGGTGGTTCA
TCTTCTGTTCGTGTGGGACGTGAGTTTCTGGTGGTGTTCGTCTGTTGATCGTGAAGCA
TGATCATAACATATTACATCTTGATGATTTAAATTTCTTTGATTTGAACTGCTGATTTTA
ATACTGCATGATCCATTTGACGAAGCTCAATCGTCTCGAGTATATCTCTATTATCTAAC
AGTATTGAGAAAAAAGGAGTTTCACTAAAAAAGGAGTTTCACTAAAAAAGGAGTTTCACT
>G2139 Amino Acid Sequence (conserved domain in AA coordinates:14-69)
MSSTKQAKGRKTKGKQKIEMKVENYGDRLMITFSKRKTGIFKKMNLVAMCDVEVAFLIF
SQPKKPYTFAPSPMKKVADRLKNPSRQEPLEDDTRPLVEAYKKRRLHDLVKKMEALEEE
LAMDLEKLKLLKESRNEKKLDKMWNFPSSEGLSAKELQQRYQAMLELRDNLCDNMAHLRL
GKDCGSSSVRVGRRVSGGVRLFDREA*
>G2343 (1..1113)
ATGGGTCATCACTCATGCTGCAACCAGCAAAAGGTGAAGAGAGGGCTTTGGTCACCGGAA
GAAGATGAGAAGCTTATTAGATATATCACAACCTCATGGCTATGGATGTTGGAGTGAAGTC
CCTGAAAAAGCAGGGCTTCAAAGATGTGGAAAAAGTTGTAGATTGCGATGGATAAACTAT
CTTCGACCTGATATCAGGAGAGGAAGGTTCTCTCCAGAAGAAGAGAAATGATCATAAGC
CTTCATGGAGTTGTGGGAAACAGGTGGGCTCATATAGCTAGTCATTTACCGGGAAGAACA
GATAACGAGATTAAAACTATTGGAATTCATGGATTAAAGAAAAAGATACGAAAAACCGCAC
CATCATTACAGTCGCTCATCAACCGTCAGTAACCTACTGTGACATTGAATGCGGACACTACA
TCGATTGCCACTACCATCGAGGCTCTACCACCACAACATCGACTATCGATAACTTACAT
TTTGACGGTTTCACTGATTCTCCTAACCAATTAAATTTACCAATGATCAAGAACTAAT
ATAAAGATTCAAGAACTTTTTTCTCCATAAACCTCCTCTCTTCATGGTAGACACAACA
CTTCCTATCCTAGAAGGAATGTTCTCTGAAAACATCATCAAAACAATAACAAGAACTAAT
GATCATGATGACACGCAAGAGGAGGAAGAGAAAATGTTTGTGAACAAGCATTTCTAACA
ACTAACACGGAAGAATGGGATATGAATCTTCGTCAGCAAGAGCCGTTTCAAGTTCCTACA
CTGGCGTCACATGTGTTCAACAACCTCTCCAATTCAAATATTGACACGGTTATAAGTTAT
AATCTACCGGCGCTAATAGAGGGAAATGTGATAACATCGTCCATAATGAAAACAGCAAT
GTCCAAGATGGAGAAATGGCGTCCACATTCGAATGTTTAAAGAGGCAAGAACTAAGCTAT
GATCAATGGGACGATTCACAACAATGCTCTAACTTTTTCTTTGGGACAACCTTAATATA
AACGTGGAAGGTTCACTCTCTGTTGGAACCAAGACCCATCAATGAATTTGGGATCATCT
GCCTTATCTTCTCTTCCCTTCTTCGTTTAA
>G2343 Amino Acid Sequence (domain in AA coordinates: 14-116)
MGHHSNCCNQKVKRWLSPEEDEKLIRYITTHGYGCWSEVPEKAGLQRCGKSCRLRWINY
LRPDIRRRGRFSPREEKLIIISLHGVGNRWAHIAHLPGRTDNEIKNYWNSWIKKIRKPH
HHYSRHQPSVTFTVLNADTTISIATTIEASTTTSTIDNLHFDGFTDSPNQLNFTNDQETN
IKIQETFFSHKPPFLMVDTTLPILGFMFSENIITNNNNKNDHDDTQRGRENVCQAFLT
TNTEEWDMLNRQOEPFQVPTLASHVFNNSSNSNIDTVISYNLPALIEGNVDNI VHNENSN
VQDGEMASTFECLKRQELSYDQWDSQQCSNFFFWNDNLNINVEGSSLVGNQDPSMNLGSS

ALSSSFPSF*

>G265 (280..1317)

CTTTGGTCTTGAAGCCAAATCAAACCTTTCCTTCAATCCTCAAATTTTCGAAAATTTTC
TCTTTTGCCTTACGTTCTCTCAATCTTATTTGTAAGAAAGTTTGTTCCTTTAATCAATC
AAATCAAAGAGACTTTTGAAGATTGTTTCCCAATTTGCGTCAATCGGGATCGAGTCAAAT
CTGAAATCTTCTCCACTCATCATCTGACTATAAGACTTAATCAAGGGACTTTTGTTCGG
GTTTGGTTTTAAACGTCTTGGATTGGAAGTGGTTAAGGTATGGATGAAAATAATGGAGGT
TCAAGCTCAGTTCCACCTTTCCTTACTAAAACATATGAAATGGTTGATGATTCTTCTTCT
GACTCGGTCTGCTTGGAGCGAAAACAACAAAAGCTTCATCGTCAAGAATCCAGCAGAG
TTTTCAAGAGACCTTCTTCCGAGATTCTTCAAGCATAAGAATTTCTCAAGTTTCATCCGT
CAGCTTAATACATATGGTTTTTCGAAAAGTAGATCCTGAGAAATGGGAATCTTGAATGAT
GATTTTGTAGAGGTGACCTTACCTTATGAAGAACATTCATAGACGAAAACCGGTTTCAT
AGCCACTCGTTAGTGAATCTACAAGCGCAAAATCCTTTGACGGAATCAGAAAGACGGAGC
ATGGAGGATCAGATAGAAAGACTGAAAATGAGAAAAGAGGCCCTTCTTGCAGGATTACAG
AACCAAGAGCAAGAACGGAAGAGTTTGGAGCTGCAAGTAACGACATTGAAAGATCGGTTA
CAACATATGGAACAACATCAGAAATCAATAGTGGCATATGTTTACAGGTTTTGGGAAAA
CCAGGACTTTTCACTAAACCTCGAAAACCATGAGAGAAGAAAAGAAGATTTCAAGAGAAC
TCTCTTCTCCCAAGCAGTTTACACATAGAACAGGTCGAAAAGTTAGAATCTTCGCTAACG
TTTTGGGAGAAATCTTGTATCGGAATCATGCGAGAAGAGCGGTTTGCAGTCATCAAGCATG
GATCATGATGCAGCTGAGTCAAGTCTAAGTATTGGCGATACACGACCCAAATCATCGAAG
ATTGATATGAATCAGAGCCGCCGTTACCGTTACTGCGCTGCTCCAAAACAGGCGTT
AACGATGACTTTTGGGAACAATGTTTACAGAGAACCCTGGATCAACCGAGCAACAAGAA
GTTTCAGTCAGAGAGAAGAGATGTCGGTAATGATAATAATGGTAATAAGATTGGAAATCAA
AGGACGTATTGGTGAATTCAGGGAATGTAAATAACATTACAGAGAAAGCTTCTTGACAT
GAATGAGGTTTTTGTAAAATAGTTTTCTTTTGGTTTCACTGAGATTATTGTATGTGTTCA
TTATTTATTACTCTGTTTCTGTAAAACAAATCTCTCTATGTTTGGGAGGAGTGACA
TAAATGCATATGCAGAATTGGTTTTCAAAA

>G265 Amino Acid Sequence (domain in AA coordinates: 11-105)
MDENNNGSSSLPFLTKTYEMVDDSSSDSVVAWSENKSFIVKNPAEFSRDLLPRFFKHK
NFSSFIRQLNTYGFRLKVDPEKWEFLNDDFVRGRPYLMKNIHRRKPVHSHSLVNLQAQNP
TESERRSMEDQIERLKNEKEGLLAELQNEQERKEFELQVTTLKDRLOHMEQHQSIVAY
VSQVLGKPGLSLNLNHHERRRRFQENSLPSSSHIEQVEKLESSLTFWENLVSESCEKS
GLQSSMDHDAESSLSIGDTRPKSSKIDMNSPPVTVTAPAPKTGVNDDFWEQCLTENP
GSTEQQEVQSERRDVGNDNNGNKNIGNQRTYWWNSGNVNNITEKAS*

>G2792 (1..960)

ATGGATCATCATCATACATAGCATCAAGAAATTCATCAACAACATCAGAATTACCATCA
TTCGAGCCAGCGTGCCATAACGGTAATGGTAACGGTTGGATCTATGACCCAAATCAAGTT
AGTACGATCAAGTAGTGACCAACGGCTGTCAAAGTTGACGGATCTTGTAGGCAAGCAC
TGGTCAATTGCACCACGAATAATCCCGACATGAACCATAACCTTCATCATCACTTCGAT
CATGATCATTCTCAAAACGACGACATTTCTATGTACAGACAAGCCTTGGAGGTGAAAAAT
GAGGAAGATCTTTGTTACAATAATGGCTCAAGTGGTGGTGGTTCTTGTTCATGATCCT
ATAGAAAGTTCTAGAAGTTTCCTTGATATAAGGTTAAGTAGGCCATTAACGGATATTAAT
CCGTCAATTTAAGCCATGCTTTAAGGCCTTAAACGTATCCGAGTTTAAACAAGAAAGACAT
CAAACGGCATCTCTGGCAGCAGTGAGACTGGGAACAACAACGCTGGAAAAAAGAAGAGA
TGTGAAGAAATTTCCGATGAGGTTTCAAAGAAGGCCAAGTGCAGTGAGGGCTCTACACTT
TCGCAGAGAAGGAACCTACCAAGCCAACTTCGAGACAAGATCACGACTCTACAGCAA
ATTGTGTCTCCCTTTGGAAAGACTGATACTGCTTCTGTGCTTCAAGAGGCCATCACTTAC
ATAAATTTTTATCAAGAGCAAGTTAAGCTGCTAAGCACTCCTTATATGAAGAATTCATCA
ATGAAGGATCCATGGGGGGGATGGGACAGAGAAGATCACAACAAAAGGGGACCGAAGCAT
CTTGATCTAAGGAGTAGAGGGCTTTGTTTGGTTCTATTTTATATACCCCAATCGCATAC
CGCGATAACAGTGCAACTGACTACTGGAATCCCACGTATAGAGGTTCTTTGTATCGTTAG
>G2792 Amino Acid Sequence (domain in AA coordinates: 190-258)
MDHHHHIASRNSSTSELPSFEPACHNGNGNGWIYDPNQVRYDQSSDQRLSKLTDLVGKH
WSIAPPNNPDMHNLHHHFDHDSQNDDISMYRQALEVKNEEDLCYNNNGSSGGGSLFHDP
IESSRSPFLDIRSLPLTDINPSFKPCFKALNVSEFNKKEHQATSLAAVRLGTTNAGKKKR
CEEISDEVSKKAKCSEGSTLSPEKELPKAKLRDKITTLQQIVSPFGKTDASVLQEAIT
INFYQEQVKLLSTPYMKNSSMKDPWGGWDREDHNRGPKHLDLRSRGLCLVPISYTPAIY

RDNSATDYWNPTYRGSLYR*

>G2830 (1..903)

ATGTCTTCCATCCCAAATAGGTTCAATATTTATGGTGGTGATACCACAAACCATCGTGAA
TCGCTTCCCATCGAAATGAATCACAACCTCTCGAATGGTTCGATCCATGTTCAATTACATCT
GATCGCATGAATCATAGAGATTTGTTTCTTCTCCTCCTTCTTTCTCTTATCAAAAT
TCACATATCTCTTCATCTTCTGTTGGGTTTAATAATTCACATATGACTTATCATATGCTG
AAAAGAAATTATGATTCTGTTTCCCGTGCTGATTATTTCTCTACTAAAGATCATTTCTCAT
TTTACTCAAGTATCTTTCACTCAAACCATCACAAATAAGTATACTACTATTGTTCTCTTCC
AATATATTTGACACTGTTCACTATGATATTGGTTCGTGTCAAACGTGCCATAGATTTTAGA
AATATTTGGAATCCTAAATCTCATCTTCCAAAAAAATTTAATAGGCAATGCGAGATTTTG
AATCCTACCCCTCTTAATATCGTCTTCCGCACCAGGATTCAGCTGATCGTCAACATTTA
GACATTATTTTCTCGTCATCAAAGCACAACCATGTTTTCCAAGATGGTCGATCCTTGAAG
AAAATTTCCGAACCAACCAATCTGTTTGAAAAATCTAATTCTTATGATTCTCAAGAAGAT
GAGAAAATCGATGCTTATCAATATGATGGTCGTACACATAGTCTACCGTATACGAAATAC
GGTCCATATACATGTCCCAGGTGTAACGGTGTGTTTGATACTTCTCAAAAATTTGCTGCA
CATATGTTATCTCACTACAATAATGAGACGGACAAAGAAAGAGACCAAAGATTTCTGTGCA
AGAAATAAAAAACGATATCGTAAGTTTATGGACAGTCTTAAATATCAAAACAGAAGATA
TGA

>G2830 Amino Acid Sequence (domain in AA coordinates:245-266)

MSSIPNRFNIYGGDTTNHRESLPIEMNHNSRMVRSMFITSDRMNHRDLFSSPPSFSSYQN
SHISSSVGFNNSHMTYHMLKRNYSVSRADYFSTKDHSHFTQVSFTQTI TNKYTTIVPS
NIFDTVHYDIGRVKRAIDFRNIWNP KSHLPKKFNRQCEILNPTPLNIVFP HQDSADRQHL
DIIFSSSKHNHVFDQGRSLKKISEPTNLF EKSNYSYDSQEDEKIDAYQYDGRTHSLPYTKY
GPYTCPRCNGVFDTSQKFAAHMLSHYNNETDKERDQRFARNKKRYRKFMDSLKISKQKI
*

>G286 (94..2454)

TGCAATTTCTCTCGACCAAAACCTAATTTCAAGGTTTGGGGTTTTCTTCTTTCACTGTC
AATTTTGATGAACTTGTGATTAGTATTAGAAATGAATGCTAATGAGCAAACCTCGATCC
GCCAATGGCATTGGCAATGGCAATGGTGAGTCTATTTCCCGGATTCAGATGACTTACGG
TGCAAGAGATCGGATGGTAAACAGTGGAGATGCACTGCAATGTCCATGGCTGATAAGACT
GTTTGTGAGAAGCACTACATCCAAGCAAAGAAAGCGGGCGGCTAATTCTGCTTTCAAGGCG
AACCAGAAGAAAGCGAAAAGGCGATCATCGTTAGGCGAAACAGATACGTATTCGGAAGGG
AAGATGGATGATTTGAGTTACCAAGTACCAGCATTGACCACTATAATAACGGTCTTGCC
TCTGCTTCCAAGAGTAATGGTAGACTAGAGAAGAGACATAATAAAAGCCTGATGCGGTAC
TCGCCCGAGACACCGATGATGAGGAGTTTCTCTCCACGTGTTGCAAGTGGATTTGAATGAT
GACTTGGGTAGAGATGTTGTAATGTTTGAAGAGGGCTACAGATCTTATAGGACACCACCA
TCTGTTGCTGTTATGGATCCGACACGAAACAGATCACACCAAAGCAGTCCATGGA
TACTCAGCAGCAAGCACAGATGTGTCTGCAAGTCTTTGGGGGAAATCTGCCATCAATGC
CAGAGAAAAGATAGAGAGAGAATCATTTCTTGCCTCAAATGCAATCAAAGAGCCTTCTGC
CACAATTGTCTATCGGCAAGGTACTCGGAGATATCACTTGAAGAAGTGCAGAAAGTTTGC
CCTGCATGTCTGCGGCTTGTGTGATTGCAAACTTGCCTGCGTTCAGATAATACAATAAAG
GTTCCGGATCCGGGAAATACCCGTTTTGGACAAGTTGCAGTATCTTTATCGTCTATTATCA
GCTGTCTTACCAGTCATAAAGCAGATCCATCTTGAACAATGTATGGAAGTTGAAC TAGAG
AAGAGGCTTCTTGAAGTTGAGATTGATCTTGTGAGGGCAAGATTGAAAGCAGATGAGCAG
ATGTGCTGCAACGTGTGTGCGATACCAAGTTGTTGACTACTACCGTCACTGTCCGAAGTGC
TCATATGACCTTTGCTGAGATGCTGTCAAGATCTACGGGAAGAGTCTTCAAGTACGAT
AGTGGGACTAACCAAAACGTACAAGATAGAAAAGGAGCTCCCAACTAAAACCTAACTTT
TCATACAAGTTTCTGAGTGGGAAGCCAACGGTGTATGGGAGCATCCCTTGCCCTCCTAAG
GAGTATGGAGGCTGCGGTTTCACTTCTTTGAATCTTGCCCGCATTTTCAAGATGAATTGG
GTTGCAAAGCTTGTGAAAAATGCTGAGGAGATTGTTAGTGGCTGCAAAATTATCTGATCTT
CTGAACCTGATATGTGTGATTCAAGATTCTGCAAAATTTGCTGAGAGAGAAGAGAGCGGT
GACAACTACGTGTACAGCCCGTCTGCTTGAACGATTAAAACCTGATGGAGTAGCTAAGTTT
GAGCAACAATGGGCAGAGGGTCTGCTTGTACTGTGAAAATGGTACTTGATGACTCATCT
TGCTCTAGATGGGATCCTGAGACTATTTGGAGGGATATAGACGAGCTTTCCGACGAGAAA
CTGAGAGAACATGATCCATTCTTGAAGGCCATTAATTGCTTGGATGGTTTAGAGGTTGAT
GTAAGACTTGGGGAGTTTACAAGAGCATATAAAGATGGAAGAAGCAAGAGACAGGTCTT
CCGCTATTGTGGAAGTTAAAGGACTGGCCGAGCCCAAGTGTCTCCGAGGAGTTTCAATTTT

TACCAAAGACCTGAGTTTATCAGAAGTTTTCCGTTTCTCGAGTACATTCATCCCCGGTTA
GGCCTTCTGAATGTTGCAGCCAAGTTACCTCATTACTCGCTCCAAAACGATTACAGTCCA
AAGATTATGTGTCTTGTGGGACGTACCAAGAAATCAGTGTGGCGATTCAATTGACTGGT
ATTCATAACAATGCGTGACATGGTATACCTATTGGTGACACAGTCTGAAGAAACAACA
TTCGAAAAGGGTGAGAAAAACAAACCTGTTCCAGAGGAACCTGACCAGAAGATGAGCGAA
AATGAGTCACTTCTTAGCCCTGAGCAGAAATTAAGGGACGGAGAGTTACATGATCTATCA
CTTGGTGAAGCCAGTATGGAGAAGAATGAACCTGAGTTGGCGTTGACTGTGAATCCAGAG
AACTTAACGGAACCGGTGACAACATGGAATCTTCTTGACATCTTCATGTGCAGGAGGA
GCCCAGTGGGATGTCTTTCGACGCCAAGACGTCCCAAAGTTGTCCGGGTATTTGCAGAGA
ACATTCCAGAAGCCTGATAATATCCAGACTGATTTTGTAAGCCGTACCTGCTAATTCAA
TAAATGAAGTGTGTAAAGTCTTGTATGTGGAATGATTGAGTTTCTAGTTTGTCTTACTCT
GGTTTCAGGTGTACGCCCCGTGTATGAAGGATTGTCTTTAAATGAACACCACAAGAGAC
AATAAGAGACGAGTTTGGAGTTGAGCCATGGACATTTGAGCAACATCGTGGTGAGGCTA
TCTTCATTCCGGCTGGATGTCGTTCCAAATCACTAATCTTCAGTCGAATATTCAGGTGG
CACTTGACTTCTTGTGCCCTGAAAGCGTTGGAGAGTCAGCAAGACTAGCTGAAGAAATCC
GGTGTTTACCAAACGACCACGAGGCAAAACCTCAGATTCTAGAGATTGGAAAGATATCAT
TATACGCAGCTAGCTCAGCCATTAAAGAGGTTTCAAGAACTGGTCTTGGATCCAAAGTTTG
GAGCAGAGCTTGGATTTGAAGACTCTAACTTAACCAAAGCAGTCTCTCAAACTTAGACG
AGGCAACCAAGCGGCC

>G286 Amino Acid Sequence (domain in AA coordinates: TBD)
MNANEQTRSANGIGNNGESIPGIPDDLRCRSDGKQWRCTAMSMADKTVCCKHYIQAKK
RAANSAFRANQKAKRRSSLGETDTYSEGMDDFELPVTSIDHYNNGLASASKSNGRLEK
RHNKSLMRYSPETPMR\$FSPRVAVDLNDLGRDVVMFEEGYRSYRTPPSVAVMDPTRNR
SHQSTSPMEYSAASTDVSAESLGEICHQCQRKDRERIISCLKCNQRAFCHNCL\$ARYSEI
SLEEVEKVC\$PACRGLCDCKSLRSDNTIKVRIREIPVL\$DKLQYLYRLLSAVLPVIKQIHL
EQCMEVELEKRLLEVEIDLVRARLKADEQMCNCVCRIPVVDYRHC\$PNCSYDLCLRCCQD
LREESSVTISGNTQNVQDRKGAPKLKLNFSYK\$PEWEANGDGSIPCP\$KEYGGCGSHSLN
LARIFKMNWVAKLVKNAEEIVSGCKLSDLLNPD\$MCD\$R\$FCKFAEREESGDNYVYSP\$LET
IKTDGVAKFEQQWAEGR\$LVTKMVLDDSSCSRWD\$PETIWRDIDEL\$DEKLR\$HDPFLKAI
NCLDGL\$EVDVRLGEFTRAYKDGKNQETGLPL\$LLWKLKDW\$PSP\$ASEEFIFYQRPEFIR\$FP
FLEYIHPRLGLLNVA\$AKLPHYSLQND\$SGPKIYVSCGYQEISAG\$SLTGIHYNMRDMVYL
LVHTSEETTFERVRKTKPVPEEPDQKMS\$ENESLLSPEQKLRD\$GELHDL\$SLGEASMEKNEP
ELALT\$VN\$PENLTENGDNMESSCTSSCAGGAQWDVFR\$QDVPKLSGYLQRTFQKPDNIQTD
FV\$RTC*

>G291 (124..1197)
CAAGAACCCAAAGATCTCTCTATTTGTTTGCCTTCTTCTTTCTTTCTGACTCAAACCC
TCAAATCAATTCTCGCGATTAAAGCAAACCTAGATTTATCTACTCTTCGAAGTCGATT
TCAATGGAAGGTTCCCTCGTCAGCCATCGCGAGGAAGACATGGGAGCTAGAGAACAACATT
CTCCCAGTGGAAACCAACCGATT\$CAGCCTCCGACAGTATATTCCACTACGACGACGCTTCA
CAAGCCAAAATCCAGCAGGAGAAGCCATGGGCTCCGATCCTAACTACTTCAAGCGCGTT
CACATCTCAGCCCTTGCTCTTCTCAAGATGGTGGTT\$CACGCTCGCTCCGGTGGCACAATC
GAGATCATGGGTCTTATGCAGGGTAAAACCGAGGGTGATACAATCATCGTTATGGATGCT
TTTGCTTTGCGCTGTTGAAGGTACTGAGACTAGGGTTAATGCTCAGTCTGATGCCTATGAG
TATATGGTTGAATACTCTCAGACCAGCAAGCTGGCTGGGAGGTTGGAGAACGTTGTTGGA
TGGTATCACTCTCACCTGGGTATGGATGTTGGCTCTCGGGTATTGATGTTTCGACACAG
ATGCTTAACCAACAGTATCAGGAGCCATTCTTAGCTGTTGTTATTGATCCAACAAGGACT
GTTTCGGCTGGTAAAGGTTGAGATTGGGGCATT\$CAGAACATATCCAGAGGGACATAAGATC
TCGGATGATCATGTTTCTGAGTATCAGACTATCCCTCTTAACAAGATTGAGGACTTTGGT
GTACATTGCAAACGACTACTACTCATTTGGACATCACTTATTTCAAGTCATCTCTCGATAGT
CACCTTCTGGATCTCCTTTGGAACAAGTACTGGGTGAACACTCTTTCTTCTTCCCCACTG
TTGGGCAATGGAGACTATGTTGCCGGGCAAATATCAGACTTGGCTGAGAAGCTCGAGCAA
GCGGAGAGTCAGCTCGCTAACTCCCGGTATGGAGGAATTGCGCCAGCCGGTACCAAAGG
AGGAAAGAGGATGAGCCTCAACTCGGAAGATAACTCGGGATAGTGCAAAGATAACTGTC
GAGCAGGTCCATGGACTAATGT\$CACAGGTTATCAAAGACATCTTGTTC\$AATTCGCTCGT
CAGTCCAAGAAGTCTGCTGACGACTCATCAGATCCAGAGCCCATGATTACATCGTGAAGT
TGGTCTATTCTTTTGT\$TTTTTGGCTGCGGAAATTGACTATCGGTTTGACCCGGTTTATGA
GGCAATGCCCATGTTCCCTATATCTCTAGTGTAGTATCTGCTTCAGACAAAGATCTTTG

GGTTATTAAATGACATTAACATAAAAAAAA

>G291 Amino Acid Sequence (domain in AA coordinates: 132-160)
MEGSSSAIARKTWELENNILPVEPTDSASDSIFHYDDASQAKIQEKPWASDPNYFKRVH
ISALALLKMOVHARSAGGTIEIMGLMQGKTEGDTIIVMDAFALPVEGTETRVNAQSDAYEY
MVEYSQTSKLAGRLNVVGVYHSHPGYGCWLSGIDVSTQMLNQYQEPFLAVVIDPRTV
SAGKVEIGAFRTYPEGHKISDDHVSEYQTIPLNKIEDFGVHCKQYYSLDITYFKSSLDSE
LLDLLWNKYVNTLSSPLLNGNDYVAGQISDLAEKLEQAESQLANSRYGGIAPAGHQRR
KEDEPQLAKITRDSAKITVEQVHGLMSQVIKDILFNSARQSKKSADDSSDPEPMITS*

>G427 (49..1230)

TTTCCCTCTCCGAAACAGAAATTCAAAAACAAATTCACACGAAAACGATGGCGTTTCAT
AACAACTACTTTAATCATTTTACCCGACCAACAACATCAGCCTCCTCCTCCGCGCAA
CAGCAGCAGCAACAACATTTTCAAGAATCAGCAGCCCCCTAATTGGCTCCTCCGCTCCGAC
AACAACTTCTCAATCTCCACACAGCTGCCACAGCCGCGCTACAAGCTCCGATTCTCCT
TCTTCCGCGCGCGCTAACACAGTGGCTCTCACGATCCTCATCCTTCTCCACGAGGCAAC
ACCGCAAACAACAACAACAACGAAACATCCGGTGACGTCATCGAAGACGTTCCCGCGGA
GAGGAGTCAATGATCGGAGAGAAGAAGGAGGCGGAGAGGTGGCAGAATGCGAGACACAAG
GCGGAGATACTGTCTCATCCACTATACGAGCAACTTTTGTGGGCACACGTGGCGTGCCTG
AGGATCGCAACCCCGGTGGATCAGCTTCCGAGGATAGACGCACAGCTTGCTCAGTCTCAA
AACGTCGTGGCTAAGTACTCAACTTTAGAAGCCGCTCAAGGACTCCTCGCCGCGATGAC
AAGGAGCTTGACCACTTCATGACGCATTATGTACTATTGCTTTGCTCTTTCAAAGAACA
CTGCAACAGCATGTTCTGTTCATGCAATGGAAGCTGTTATGGCCTGTTGGGAGATTGAA
CAGTCGCTTCAAAGTTTACAGGAGTATCTCCTGGTGAAGGCACAGGAGCAACAATGTCT
GAGGATGAAGATGAGCAAGTAGAGAGTGATGCTCATTTGTTTGTATGGAAGCTTAGATGGG
TTAGGGTTTGGTCTCTAGTTCCCACTGAGAGCGAGAGATCTTTGATGGAACGAGTCAGA
CAAGAACTCAAACATGAACCTCAAGCAGGTTTACAAGGAGAAAATTGTGGACATAAGAGAG
GAGATACTGAGGAAGAGAAGAGCTGGAAAATTACCAGGAGACACCACCTCTGTCTCAAA
TCATGGTGGCAATCTCATTCTAAGTGGCCTTACCCTACTGAGGAAGATAAGGCGAGGTTG
GTGCAGGAGACGGGTTTGCAGCTCAAACAGATAAACAATTGGTTCATCAATCAAAGAAAG
AGGAATTGGCATAGCAATCCATCTTCTTCTACCGTCTCAAAGAATAAACGCCGAAGCAAT
GCAGGTGAAAACAGCGGAAGAGACCGTTGAGATCAAGCTTGATGTAGAGATCCAAAAGC
TTTATAGAAAGGTGGAGGCATGAAGACAAAGAATTCTTACACAACAACGTTAGGACGTAA
TTTTGTGCCAGTACATGGTATGGCTTTTCATATTTGGTAATGATTAGGGCCACACAAAATT
AAACCCCAAAGCATGATTGTAAATATGAGGTTTTAGATGGACTTTATGATAGGATCGTCA
GTCTTCACTGCCATCTCCATTCTCCACCATCAATCCATCATTTATATCTTGTGAAAAAAA
A

>G427 Amino Acid Sequence (domain in AA coordinates: 307-370)
MAFHNNHFNHFTDQQQHQPPIPPPPQQQQQHPQESAPPNWLLRSDNNFLNLHTAATAAATS
SDSPSSAAANQWLSRSSSFLQRGNTANNNNNETSGDVIEDVPGGEESMIGEKKEAERWQN
ARHKAELSHPLYEQLLSAHVACLRIATPVDQLPRIDAQLAQSQNVVAKYSTLEAAQGLL
AGDDKELDHFMTHYVLLLCFKEQLQHVHVHAMEAVMACWEIEQSLQSFTGVSPEGGTG
ATMSEDEDEQVESDAHLFDGSLDGLGFGPLVPTESERSLMERVRQELKHELKQGYKEKIV
DIREEILRKRRAGKLPDGTTSVLKSWWQSHSKWPYPTEEDKARLVQETGLQLKQINNWF
NQQRKNWHSNPSSSTVSKNRRSNAGENSGRDR*

>G509 (122..1054)

CTTCCTCCTTTGCTAATAAACTTTTCTTTGAACCTTACACGCCTTGTTGATATTACTCTC
TTAAATATATATTTTCGTACATTAACACAGACATATATAAGCTAAAGATTTCTTCACGT
AATGGGTTTGAAGATATTGGGTCCAAATTGCCACCGGGTTTCGATTTTCATCCAAGTGA
TGAAGAGTTGGTTTTCATTATCTTTGCAACAAGATTAGGGCCAAATCTGATCATGGTGA
TGTTGATGATGATGATGATGATGTTGATGAAGCTTTGAAGGGTTCTACTGATCTTGTGGA
GATTGACTTGATATCTGTGAGCCATGGGAGCTTCTGATGTGGCAAAGTTAAACGCAAA
GGAATGGTACTTCTTCAGTTTCCGTGATCGAAAGTATGCTACTGGATATCGCACGAACAG
AGCGACAGTAAGCGGATACTGGAAGCAACAGGAAAAGATCGAACGGTGATGGATCCACG
TACAAGGCAAAATGGTAGGGATGAGAAAAACACTAGTGTTCTACAGAAACAGAGCACCAAA
TGGGATCAAAACTACTTTGGATCATGCACGAGTTCCGTCTTGAGTGTCTTAACATCCACA
TAAGGAAGACTGGGTCTTGTGTCAGAGTGTTCACAAAGGCAGAGACTCATCGCTACAAGA
CAATAATTATTATAACAATGATAATCAGACGCAAGGCTTGAAGTTAATGACGCTCCGGA
TCTTAATTACAACAATCAGTTGCCACCTTTGCTATCATCCCCTCCTCATAATCATCAACA

TGAGAAGATGAAAATCCAAGTTTGTGATCAGTGGGAGCAGCTAATGAAGCAGCCTTCAAG
GACCACCGGCCACCCCTATCATCACCATTGTTCATCATCAAACCATAGCATGTGGTTGGGA
GCAGATGATGATCGGTTTCGCTGTTCATCACCTTCGAGTCATGGCCCTGATCAGAGTCCCTT
TGCTAAATTTGCTTTACCGTCGACAATAACAACAGTGTCAACATCAGTGGTGATCATCAT
CAGAAATTATGAGAAGATTTTGTGTTCATCACTAGACATGACGAGTTTGGATCAGACAAG
ACATGTATGGGATCATCATCGGATGGTGGTATGGTCTCTGATCTTCACATGGAATGTGGT
GGATTGAGTTTGTAGACCGAAAATATCCTCGCTTCCAATGAACATAATTCAAGGGGTTTC
GCCAATTTGTTGATTCTGTGAATTATACAAACATTTATCTATAGATTTATCACATTATCA
AACATGTAAGTTGTGTGGCATTGTGGGTATAGGGTTGTGTGATTCTAGGTTTTTAGGACG
ATGTATGTTGTTATATTTAGCGTGTTTTTAGGATTTATTCTCATTTTAAATATATGAA
AACCCTACTATGAATACAATTAGTTTCTTTGTTGTAAATAATATTTAGATTATCAA
AAAAAAAAAAAAA

>G509 Amino Acid Sequence (domain in AA coordinates: 13-169)
MGLKDIGSKLPPGFRFHPSPDEELVCHYLCNKIRAKSDHGDVDDDDDDVDEALKGSTD LVE
IDLHICEPWELPDVAKLNAKEWYFFSFRDRKYATGYRTNRA TVSGYWKATGKDRTVMDPR
TRQLVGMKRLTVFYRNRA PNGIKTTWIMHEFRLECPNIPHKEDWVLCRVFNKGRDSSLQD
NNYYNNDNQTRLEVNDAPDLNYYNQLPPLLSSPPHNHQHEKMKIQVCDQWEQLMKQPSR
TTGHPYHHHCHHQTIA CGWEQMMIGSLSSPSHGP DHESFAKFALPSTITT VSTS VVII
RIMRRFCCHH*

>G519 (85..894)

CACAAAGATCCTCCGATTGGAAGGTTTATAAAAACTCAAATCGAATCTTATCCACAAGA
AAACAACAAGGTACTTTTCCAAAAATGAAGGCGGAGTTGAATTTGCCGGCGGGATTCCGA
TTTCATCCGACGAGCAAGAGCTTGTCAAGTTCATCTTTGCCGGAGATGTGCGTCAGAA
CCGATTAACGTTCCGGTTATCGCAGAGATTGACTTGTACAAATCAATCCATGGGAGCTT
CCAGAAATGGCGTTGTACGGTGAGAAAGAATGGTACTTCTTCTCGCATAGAGACCGGAAA
TACCCAAACGGGTCGAGACCAACCGGGCAGCTGGAACCGGTTATTGGAAAGCGACTGGA
GCTGATAAACCGATCGGAAAACCGAAGACGTTAGGGATTAAGAAAGCACTCGTCTTCTAC
GCAGGAAAAGCTCCGAAAGGGATTAAAACGAATTGGATTATGCACGAGTATCGTCTCGCT
AATGTCGATCGATCTGCTTCTACCAACAAGAACAACCTTAAGACTTGATGATTGGGTT
TTGTGTCGATATCAATAAGAAAGGAACAATGGAGAAGTATTTACCGGCGGCGGCTGAG
AAACCGACGGAAGATGAGTACGTCGGACTCAAGATGCTCAAGTACGATGATTTACCG
GACGTCACGTGTTCTGATAACTGGGAGGTTGAGAGTGAGCCCAAATGGATTAATCTGGAA
GACGCGTTAGAGGCATTTAATGATGACACGTCCATGTTTAGTTCCATTGGTTTGTGCAA
AATGACGCCTTTGTTCTCAGTTTCAGTACCAAGTCCCTCGATTTCGTCGATTGTTTTAG
GACCGGTTTCGAGCAGAAACCGTTCTTGAATTGGAATTTTGCTCCTCAAGGGTAAAAATAA
TCGGCAAAAAGTTGAAGCTTTTCAGAGTCTTCGATCACCGGCATTGTGTCGGATCCTGAC
CCGGAGACCAAGTCGGGTACATACGATTACATAATCGGGTTATTGAGATTCCACATTTGG
ATTTCCGAGACTAACCAACTTAACGGATTCTGGGGTAATTGGGGGTTTTCACAGGTGA
ATCACACTGAGTCAGCAAGTTTCGATTTTTTGGTTTGTGTTTGTAAATGATTGATTAAATG
TCTAAAGATATACGAAGTAGATTGAGAAGAACTGTAAAAGCAATTGTGACCACCCGTTA
TGAATCATAAATATATTCAATGAAGCATGAGCTTATTTTTTTTTTAAAAA

>G519 Amino Acid Sequence (conserved domain in AA coordinates: 11-104)

MKAELNLPAGFRFHPPTDEELVKFYLCRRCASEPINVPVIAEIDLYKFN PWELPEMALYGE
KEWYFFSHRDRKYPNGSRPNRAAGTGYWKATGADKPIGPKTLGIKKALV FYAGKAPKGI
KTNWIMHEYRLANVDRSASTNKKNNRLRDDWVLCRIYNNKGTMEKYLPA AAEKPTEKMST
SDSRCSHVISP DVTCSDNWEVESEPKWINLEDALEAFNDDTSMFSSIGLLQND AFVPQF
QYQSSDFVDSFQDPFEQKPF LNWNFAPQG*

>G561 (86..1168)

AATTTGTTTTTTTTTCTTTTGTGGGTTCAATTGCAATTGTTTTCCCTGAGACTCAAGTTA
CTGTGTCATTACTCTGCATTGAGCAATGGGTAGCAACGAAGAAGGAAACCCACTAACAA
CTCTGATAAGCCATCGCAAGCTGCTGCTCCTGAGCAGAGTAATGTTTCATGTGTATCATCA
TGACTGGGCTGCTATGCAGGCATATTATGGGCCTAGAGTTGGTATACCTCAATATTACAA
CTCAAATTTGGCGCCTGGTCATGCTCCACCGCTTATATGTGGGCGTCTCCATCGCCAAT
GATGGCTCCTTATGAGAGCACCATATCCACCATTTTGGCCCTCCTGGTGGAGTTTATGCTCA
TCCTGGTGTTCAAATGGGCTCACAAACCACAAGGTCCTGTTTCTCAATCAGCATCTGGAGT
TACAACCCCTTTGACCATTGATGCACCAGCTAATTCAGCTGGAACTCAGATCATGGGTT
CATGAAAAGCTGAAAGAGTTTCATGGACTTGCAATGTCAATAAGCAATAACAAAGTTGG

GAGTGCTGAACATAGCAGCAGTGAACATAGGAGTTCTCAGAGCTCCGAGAATGATGGCTC
TAGCAATGGTAGTGTGTAATAACAACCTGGGGGAGAACAACTCTAGGAGGAAAAGAAGGCA
ACAAAGATCACCAAGCACTGGTGAAAGACCCTCATCTCAAAACAGTCTGCCTCTTAGAGG
TGAAAATGAGAAACCCGATGTGACTATGGGGACTCCTGTTATGCCACAGCAATGAGTTT
CCAAAACCTCTGCTGGCATGAACGGTGTGCCACAGCCATGGAATGAAAAAGAGGTAAACG
AGAGAAGAGAAAACAGTCAAACCGAGAATCTGCTAGGAGGTCAAGACTGAGGAAGCAGGC
TGAAACAGAACAACTATCTGTCAAAGTTGACGCATTAGTAGCTGAGAACATGTCTCTGAG
GTCTAAACTAGGCCAGCTAAACAATGAGTCTGAGAACTACGGCTGGAGAACGAAGCTAT
ATTGGATCAACTGAAAGCGCAAGCAACAGGGAAAACAGAGAACCTGATCTCTCGAGTTGA
TAAGAACAACTCTGTATCAGGTAGCAAACTGTGCAGCATCAACTGTTAAATGCAAGTCC
GATAACCGATCCTGTGCGGGCTAGCTGACCGTGGCCGCAACAATGAGAACCCGATATTTC
TTCTTTTGGGTTGTGATTGTAACCTTAAAGGAGACTTTTTTGTTTTTTATCTTAGATTGT
AGCTCTCTGCATAGTGAGCATAAATTGATGTAATATGGTTTAAAGAGATTCGGTGTCTCT
GGTGTGTGCTGCAACCACATAATTGGTGATAGATAGGTTTAGTTATATAAGCAAATGTAT
TAGAGATAAGGGGAGACATATTGATGGTCTTT

>G561 Amino Acid Sequence (domain in AA coordinates: 248-308)
MGSNEEGNPTNNSDKPSQAAAEQSNVHVYHHDWAAMQAYYGPVGIQYYNSNLAPGHA
PPPYMWASPSPPMAPYGPYPFPPGPGVYAHQVQMGSPQGPVVSQASGVTTPLTIDA
PANSAGNSDHGFMKKLKEFDGLAMSI SNKVGSAEHSSEHRSSQSENDGSSNGSDGNT
TGGEQSRKRQRSPSTGERPSSQNSLPLRGENEKPDVTMGTPVMPATMSFQNSAGMNG
VPQPWNEKEVKREKRKQSNRESARRSLRKQAE'EQLSVKVDALVAENMSLRSLKGLQNLN
ESEKLRLNEAILDQLKAQATGKTENLISRVDKNNSVSGSKTVQHQLLNASPIIDPVAAS

*

>G590 (102..1223)
TCGACAGACACTCTCCCTCTCTCCATGCCCATAAAATCTCAAAGACTGTTTAAAAA
AATGTTTTAGCTTTAACTGCTTTTTTTTGTGTGGTGTAATGATATCACAGAGAGAAG
AAAGAGAAGAGAAGAAGCAGAGAGTGATGGGAGATAAGAAATTGATTTCATCTTCTCTT
CTTCTCGGTTTACGATACTCGTATCAATCATCATCTTCATCATCCTCCGCTCTTCTCCG
ACGAAATCTCTCAGTTTCTCCGCATATTTTCGACCGTTCTTCTCCTTTACCTTCTTACT
ACTCCCCGGCGACGACTACAACGACGGCGTCTTTGATTGGTGTGCACGGGAGCGGTGACC
CACATGCAGATAACTCGAGAAGTCTCGTTTCTCATCATCCACCGTCAGATTCTGTGCTTA
TGTCGAAACGTGTGCGGAGATTTCTCTGAGGTTTTAATCGGCGGAGGATCAGGCTCAGCCG
CCGCGTGTTTTGGTTTTCTCCGTTGGTGGTAATAATAACAACGTTCAAGGAAATAGCTCTG
GGACTCGAGTATCGTCTTCTCCGTTGGAGCTAGTGGCAACGAGACAGATGAGTATGACT
GTGAAAGCGAGGAAGGAGGAGAAGCTGTAGTTGATGAAGCTCCCTCTTCCAAGTCAGGTC
CTTCTTCTCGTAGTTTCATCTAAAGATGCAGAGCTGCTGAAGTTTCATAATCTCTGTGAGA
AGAGGAGGAGAAGTAGAATTAATGAAAAAATGAAAGCTTTACAAAGTCTCATCCCTAATT
CAAATAAGACGATAAGGCTTCAATGCTTGATGAAGCCATTGAGTATCTGAAACAGCTTC
AGCTCCAAGTTCAGATGTTGACTATGAGAAATGGAATAAACTTGCATCCTTTGTGTTTAC
CTGGAACCTACATTACACCCATTGCAACTCTCTCAGATTCGACCCCTGAAGCAACCAATG
ATCCTCTGCTTAATCATACCAATCAGTTTGCTTCGACTTCTAATGCACCGGAAATGATCA
ATACTGTGGCTTCTTCATACGCTTTGGAACCTTCTATTTCGAGTCACTTTGGACCTTTCC
CTCTCCTTACTTCACCCGTGGAGATGAGTCGGGAAGGTGGGTTAACTCATCCAAGGTTGA
ACATTGGTCATTCCAACGCAACATAACCGGGGAACAAGCTCTGTTTTGATGGACAACCTG
ACCTAAAAGATCGAATTACTTGAACAGTGTCCCAACTTCGGGATCTCTATGTGTTCTTGT
TTCTTAGAACGCAAGCCATAAAGCTGTCTGAC

>G590 Amino Acid Sequence (domain in AA coordinates: 202-254)
MISQREBEREKKQVRMGDKKLISSSSSSVYDTRINHHLLHPPSSSDEISQFLRHIFDRS
SPLPSYYSPATTTTTASLIGVHSGDHPHADNSRSLVSHHPPSDSVLMSKRVGDFSEVLIG
GGSGSAAACFGFSGGNNNVQGNSSGTRVSSSSVGASNETDEYDCESEEGGEAVVDEA
PSSKSGPSSRSSSKRCRAAEVHNLSEKRRRSRINEKMKALQSLIPNSNKTDKASMLDEAI
EYLKQLQLQVQMLTMRNGINLHPLCLPGTTLHPLQLSQIRPPEATNDPLLNHTNQFASTS
NAPEMINTVASSYALEPSIRSHFGPFPPLTSPVEMSREGGLTHPRLNIGHSNANITGEQA
LFDGQPDLDKDRIT*

>G818 (65..1060)
GTATTTCTTACAATAAACGACCAAAAAGTTAATACAAGAAATAGAAACGGTGTAGGAAGC
TACTATGACGGCAATTCCAACGTCGTCGATATTGAATCTTCTTCTCTTCTGCTTTGTCA

AGAGACGGCAACGGAGACCGTCACCGTTGAAAGAGGCTCGTCTGATTTCATCTTCAAAGCC
AGACGACGTCGTTTTACTAATCAAGGAAGAGGAGGATGACGCCGTTAACTTGTCACTTGG
TTTTTGGAAATTGCACGAGATAGGTTTAATAACACCGTTCTTGAGAAAGACGTTTGAGAT
CGTCGATGACAAAGTAACAGACCCGGTTGTATCATGGAGCCCGACCCGTAAGGCTTTAT
CATTGGGATTCTTACGAGTTCTCAGAGAATCTACTTCCCAAATACTTCAAGCACAAAGAA
CTTCTCCAGTTTTATTTCGTAGCTTAACCTTACGGTTTTTAAAAAGGTCGATTAGATAG
GTGGGAATTTGCTAACGAAGGGTTTCAAGGAGGGAAGAAACATTTGCTTAAGAATCAA
GAGGAGAAGCAAAAACACTAAATGTTGTAACAAGGAAGCGAGTACCACCACGACAGAGAC
TGAGGTTGAGTCATTGAAGGAGGAACAGAGTCCAATGAGATTGGAGATGTTGAAGCTGAA
ACAACAACAAGAAGATCTCAACATCAGATGGTCACTGTGCAGGAGAAGATCCACGGAGT
TGATACCGAACACAGCATATGCTTAGTTTCTTGCAGTTGGCTAAAGATCAAAGATT
TGTAAGAGAGACTGGTGAAGAAGAGAAAGATGAAAATACAGAGAGAGCTAGAAGCAGCTGA
ATTCTGTAAGAAGCTCAAGTTGCTTCAGGATCAAGAACTCAAAAGAACTTGTAGATGT
AGAAAGAGAATTTATGGCCATGGCTGCAACAGAACACAATCCCGAGCCTGACATTTTGGT
GAACAATCAAAGCGGAATACGAGATGTCAGCTTAACCTCAGAGGACCTACTTGTGACGG
TGGCTCAATGGATGTAAATGGGAGGATAGAGATAGAGTAGAGCAAAACCGGTAACATAGC
AATAGAGAAGGTACCAATCCCAAGGCTTGAGATCCGAAT

>G818 Amino Acid Sequence (domain in AA coordinates: 70-162)
MTAIPNVVDIESSSSSLCQETATETVTVERGSSDSSSKPDDVLLIKEEDDAVNLSLGF
WKLHEIGLITPFLRKTFEIVDDKVTDPVVSWSPTKRSFIWDSYEFSENLLPKYFKHKNF
SSFIRQLNSYGFKKVDSRWEFANEGFQGGKHLKLNKRRSKNTKCCNKEASTTTTETE
VESLKEEQSPMRLEMLKLKQQQEESSQHQMVTVQEKIHGVDTEQQHMLSFFAKLAKDQRFV
ERLVKKRKMKIQRELEAAEFVKKLKLLQDQETQKNLLDVEREFMAMAATEHNPEPDILVN
NQSNGNTRCQLNSEDLLVDGGSMDVNGRIEIE*

>G849 (218..2077)

AACTCGAGAATCTTTCATTTCTTTTAAATCTTAGAATCTCGAGTTTTTGTATAAATCGAT
TCTAATTTTTCTTTGTACATTGTTTTATATATACATAAAACACAAATCGGGTATGGG
GGAATTTGGGTTTTAAGATAGCGTGATCTGTAATAATAAGTGGTTTCGCGATCGTGATCAA
GAACTGGTGGCTGATAGTGATATGCTATTTGAGAGATGGTGTTCAGAGAAAGTTAGA
TTGCCTTTCCGTGGGATTTGATTTTCCCAACATTCCAGAGCTCCTCGTTCATGCAGGAG
GAAGGTTCTAAACAAGAGGATTGATCATGATGATGATAACACTCAGATCTGTGCAATTGA
CTTACTAGCTTTGGCTGGAAGATTCTACAGGAAAGCGAGAGTTCTCTGCGTCTTCTAA
TGCAATTTGAAGAAATTAAGCAAGAGAAAGTAGAAAATTGCAAGACTATTAAATCTGAGTC
TTCTGACCAAGGAACTCTGTGTCAAAGCCTACTTATGATATCTCTACTGAGAAGTGTGT
GGTGAACAGTTGTTTTTCATTTCCGGATAGTGACGGCGTTTTGGAGCGGACTCCGATGTC
TGATTACAAGAAGATTGATGTTGATGATGATGATGATGATGATGATGATGATGATGATGAT
TGGGTTTCGAGCAAGGAGAGCAACCGATCGCGTGGGTGATGGAGGCTTAGTCACTGATAC
TTGCAACTTAGAGGATGCAACTGCGTTAGGTCTGCAGTTTCCGAAATCAGTCTGTGTGGG
TGGTGATTTAAAATCACCATCCACCTTGGATATGACCCCTAATGGTTCCTATGCTAGACA
TGGGAACCATACTAACCTAGGTAGAAAAGATGATGATGATGATGATGATGATGATGATGAT
ACTTAGCAATAAATTTAAGTCGTATAGGTCTCCAACAATTGCAAGAAATAAGAAAGTCCAT
GTCGTCCAAATACTGGAACAAGTTCAAAAGATTTTGATACAGTAGAGCTGATGTGGG
TGTGAAGACTCTTTATCGCAAAAGAAAATCATGTTATGGTTACAACGCATGGCAGCGTGA
GATCATTTATAAGAGAAGAAGATCACCTGACAGAAGCTCGGTGCTAAGTTCTGATGGAGG
ACTCAGTAGTGGAAGTGTTCGAAGTTACCCAAGAGGGAGATACAGTAAAGCTAAGCAT
TAAGTCCTTTAGGATTCAGAGCTTTTTTATTGAAGTTCAGAACTGCAACAGTAGGATC
ACTAAAGAGGACTGTGATGGAGGCTGTGAGTGTCTTACTCAGCGGAGGAATACGTGTTGG
GGTGTTAATGCATGGGAAGAAGGTTAGAGATGAAAGGAAAACCTCTGTCCAGACTGGGAT
CTCATGTGATGAAAATCTAGACAACCTTGGGTTACCTTGGAGCCTAGTCCCAGCAAAGT
TCCCTTACCTTTGTGTTCTGAAGATCCTGTGTGCCAACCGACCTACAAGTTGTCTGA
ACGGTCTGCGGCGTCTCCTATGCTAGATTCTGGAATTCACATGCAGATGACGTGATTGA
TTCAAGAAATATTGTGACAGTAACCTCGAATTAGTTCCATATCAGGGTGACATATCTGT
TGATGAACCTTACAGATTCAGAAAGGCTTGTCCCACTTCCAGAGTTGGAAGTCAAGGC
GCTTGCCATAGTTCCGTTGAACCAGAAACCTAAGCGTACTGAGCTAGCCAGAGGAGAAC
TAGGAGACCTTCTCTGTGACAGAGGTAGAAGCTCTGTACAAGCAGTTGAGGAACCTCGG
GACTGGAAGATGGCGTGATGTAAATTTGCGTGCTTTCGAGGATGCAGATCATCGGACTTA
CGTGGACTTGAAGGACAAATGGAAGACGCTAGTTCACACAGCAAGTATATCCCCACAGCA

227

CTCGAAAGCCCTAACAGTCAGGCAATCAACAACCTGCCACGTAAGCTCTCCCGACACTAAT
CATAATATCCACGTCAGCAACGTGGTCGACACTAGCTTTGTTACTAGCTGGGCGGCTTTA
GACCGCCTCGTGGCCTCGCAGCTTAACGGACCCACATCATATTCAATTACAGCCGTCAT
GAGAGCCACGTGGGCCATGATCATCTCGCTTTGCCCTCCGTCGGATCTCCGTACCCAGC
CTAAACCGGTCCGCTTCGTACCACGCCGGTTTAAACACAGGAATATACACCGGAGATGGAG
CTATGGAATACGACGACGTCGTCTCTATCGTCATCGCCTGGCCCATTTTGTACGTGTG
AATGTTTTGTGCTTGTGTTGTCTCCTTCGTCTGCAGCTTCAGTTCTGGCCGTTCCAACCA
TGGCAGAGGCAGGTTTCATTTTCATCTTCATCGCCTCAGATGCAGATCTCTCTCCATTGA
>G961 Amino Acid Sequence (conserved domain in AA coordinates: 15-140)
MSKSMSISVNGSQVPPGFRFHPTEELLQYYLRKKVNSIEIDLVDVIRVDLNLKLEPWDI
QEMCKIGTTPQNDWYFFSHKDKKYPTGTRTNRATAAGFWKATGRDKIIYSNGRRIGMRKT
LVFYKGRAPHGQKSDWIMHEYRLDDNIISPEDVTVHEVVSIGEASQDEGWVVCRIFFKKK
NLHKTLSNPVGGASLSGGGDTPKTTSSQIFNEDTLDQFLELMGRSCKEELNLDPPMKLPN
LESPNSQAINNCHVSSPDTHNHNHVSINVVDTSFVTSWAALDRLVASQLNGPTSYSITAVN
ESHVGHDLALPSVRSPPYPSLNRASASYHAGLTQEYTPEMELWNTTSSLSPPGPFCHVS
NVLLLVCLLRLQLQFWPFPWQRQVHFDLSSPQMQLSLH*

>G1465 (163..1125)

TATCCTTCGCAAGACCTTCCTCTATATAAGGAAGTTCATTTTCATTTGGAGAGGACACGC
TGACAAGCTGACTCTAGCTTATCTGGTACCGTCGACCTCATTCCTGCGTTTGATCTTCT
TTCTCTAGATCCCATATTTTTCTTGATCAATTTAGTTTCATTATGGAGGAAGATGCAGCT
TTTGATCTACTCAAAGCGAACTCTTAAACGCAGAAGACGATGCAATAATCTCACGTTAT
CTGAAGCGTATGGTCGTCAACGGAGACTCATGGCTGATCACTTCATCGAAGACGCAGAC
GTGTTCAACAAGAATCCAAATGTGGAGTTCGATGCTGAGAGCCCTAGCTTCGTGATAGTT
AAACCTCGAACAGAGGCTTGTGGTAAAACCGATGGATGTGAACTGGTTGCTGGAGGATC
ATGGGTCTGATAAACCGATAAAATCGACGGAGACTGTGAAGATTCAAGGGTTCAAGAAG
ATTCTCAAGTTCTGCCTAAAGAGGAAACCTAGAGGATACAAGAGAAGTTGGGTAATGGAA
GAGTATAGGCTTACCAATAACTTGAAGTGAAGCAAGATCATGTGATTTGCAAGATTCCG
TTTATGTTGAAGCTGAAATCAGTTTCTTGCTAGCCAAGCATTCTACACTACATCAGAA
TCACCTCCTCGAAATGAGCTGTTGCCAGCTTACGGATTCTTTCATCAGATAAGCAATTG
GAGGATGTATCTTATCCGGTGACGATAATGACTTCTGAAGGAAACGATTGGCCTAGCTAC
GTTACCAACAATGTGTATTGTCTGCATCCATTGGAGCTCGTTGATCTTCAAGATCGGATG
TTTAATGATTACGGAACCTGCATCTTCGCTAACAAAGACTTGTGGTAAAACCGATAGATGC
ATTAATGGTGGTTACTGGAAAATTTGCACCGTGATAGGCTGATCAAGTCAAAGTCCGGG
ATAGTTATTGGTTTCAAGAAGGTGTTTAAAGTTTCATGAAACGGAGAAAGAAAGATACTTC
TGTGGTGGAGAAGATGTGAAGGTAACCTGGACTCTAGAAGAGTATAGGCTTAGCGTGAAG
CAGAATAAATTCTTGTGCGTTATCAAGTTTACTTATGATAACTAAGAATCTTTCTTTGG
ATTTTATGATCATCTTAGTATCGCGACCGCTCTAGACAGGCCTCGTACCGGATCCTCTAG
CTAGAGCTTTCGTTCTGATCATCGGTTTCGACAACG

>G1465 Amino Acid Sequence (conserved domain in AA coordinates: 242-306)

MEEDAAPDLLKAELLNAEDDAIISRYLKRNVNGDSWPDHFIEDADVFNKPNVEFDAES
PSFVIVKPRTEACGKTGDCETGCWRIMGRDKPIKSTETVKIQGFKILKFCLKRKPRGYK
RSWVMEERYLTNNLNWKQDHVICKIRFMFEAEISFLLAKHFYTTSESLPRNELLPAYGFL
SSDKQLEDVSYPTIMTSEGNDWPSYVTNNVYCLHPLELVDLQDRMFNDYGTICIFANKTC
GKTDRINGGYWKILHRDLIKSKSGIVIGFKKVFKEHETEKERYFCGGEDVKVTWTLLE
YRLSVKQNKFLCVIKFTYDN*

>G425 (45..1196)

GAAAACAGTCTTCTCTCTCCGATCCCAAAAACGCAGGAAAAAATGTCGTTTAAACAGCTCCC
ACCTCCTTCTCCAAAGAAGACCTTCCTCTCCGACACTTCACCGATCAATCACAGCAACCTC
CGCCGACGCTCACTTCTCTGAAACACCTTCGCTTGTACCGCCAGTTTCTCAACCTCCCTA
CCACCCTTACCACTGCGGATTCCGATCTCGCTCCTCCGACCGCAACGGAGACAATTCGGTT
GCTGATACAAACCCACGCTGGCTCTCCTTTTCATTCCGAGATGCAAAATACTGGAGAAGTACG
TTCTGAAGTTATCGACGGAGTCAACGCCGATGGTGAACGATACTCGGCGTTGTAGGAGGT
GAAGATTGGCGGAGTGCTAGCTATAAGGCGGCGATTTTAAGACATCCGATGTACGAGCAGC
TTCTTGCGGCTCACGTGGCTTGCCTTAGGGTTGCGACTCCCGTTGACCAGATTCCGAGGATC
GATGCTCAGCTCAGTTCAGTTGCATACCGTCGCGCGAAATACTCCACTCTTGGTGTGGTTGTT
GACAACAAGGAACCTTGATCATTTTCATGTACATTATGTTGTCTTGTATGTTTCATTTAAAGAACA
ACTCCAACACCACGTTTGTGTCCATGCAATGGAAGCCATTACGGCTTGTGGGAGATTGAACA

ATCACTGCAATCCCTAACTGGAGTTTCTCCAAGTGAAAGTAATGGTAAGACAATGTCCGATGA
TGAAGATGATAATCAAGTAGAGAGCGAGGTGAACATGTTTGATGGAAGTTTGGACGGTTCAG
ATTGCTTGATGGGGTTTGGTCCTCTTGTTCACACCGAGAGAGAGATCTTTGATGGAACGTG
TGAAGAAAGAACTGAAGCATGAGCTTAAACAGGGTTTCAAAGAGAAGATTGTGGACATAAG
AGAAGAGATAATGAGGAAGAGAAGGGCGGAAAGTTGCCAGGAGATACGACTTCTGTACT
CAAAGAATGGTGGCGAACTCACTCGAAATGGCCATACCCAAGTGAAGGATAAGGCAAAA
GAACTGGAACAGCAACTCTTCCACGTCATCTACTCTACCAAGAACAAACGTAAACGGACC
GGGAAGTCGTAGGTGACATAGCGGCTAACTAGAGGATGGTTCTTTGCCATGTGAATTCTTGG
GAACCGTATATGAAAGAAACGAATCCGGTTCATGCTCGTACAGAGTGTGTTATTTGTATAGT
GGATACCGGTTAGCCTATGAAACCGGATTCTGGAGTCCAAATTGTTGTTTGTAAACGACTTAGT
AGTTTTTGGAAAGTGATCTGTTTCGTTGGTTTGGCTCTTGTAACGAACGCTTAAGCAAGTGTGGG
TTTTTTCTTGTAAGTGTCATATGTTTCGTTTGTAAATGAATGTATCAAGCAATATTTATCATAATT
AAACTAGCTTGAAATGTAAAAA

>G425 Amino Acid Sequence (domain in AA coordinates: TBD)

MSFNSSHLPPQEDLPLRHFTDQSQQPPPPQRHFSETPSLVASFLNLP TTLTADSDLAPPHR
NGDNSVADTNPRWLSFHSEMONTGEVRSEVIDGVNADGETILGVVGGEDWRSASYKAAILR
HPMYEQLLAAHVACLVRVATPVDQIPRIDAQLSQLHTVAAKYSTLGVVVDNKELDHFM SHYVVL
LCSFKEQLQHVCVHAMEAITACWEIEQSLQSLTGVSPSESNKMTSDEDDDNQVESEVNM
FDGSLDGS DCLMGFGPLVPTERERSLMERVKKELKHELKQGFKEKIVDIREEIMRKR RAGKLP
GDTTSVLKEWWRTHSKWPYPTEEDKAKLVQETGLQLKQINNWF INQRKNWNNSNSTSSTLT
KNKRKRTGKS*

>G347 (1..570)

atgaaagtagcagatatgcaggaccagctgggtgtgtcatgggttaggaatttattgatg
tatactagaggagcatctaattgtgcgttgtgcgttatgtaacactatcaacatgggtcct
cctcctcctccacctcagcagatggcacacattatatgtgggtgggttagaacaatgctt
atgtatacgcgtggggctagtagcgtaagatgctcttgctgtcaaacacgaaccttggtg
ccagcgcactccaatcaggttgcccatgctcctccagtcaggttgcgagatcaattgt
gggcattgtcgagcagacctcatgtatccttacggtgcatcatccgtcaaatgcgctgtt
tgtcaattcgtaactaacgttaatatgagcaatggaagggtacctctcccaactaaccgg
ccaaatggaacagcttggtccccctctacatcaactcaacaccacctctcagacccaa
accggttggttagaaaaccccatgtccgttgatgaaagcggaagttgggtgagcaatggt
gttggtggagtgacaactgacaaaaagtaa

>G347 Amino Acid Sequence (domain in AA coordinates: 9-39, 50-70, 80-127)

MKVADMQDLVCHGCRNLLMYPRGASNVRCALCNTINMVPPPPPPHMAHIICGGCRTML
MYTRGASSVRCSCCQTNNLVPASHNQVAHAPSSQVAQINCCHCRTTLMYPYGASSVKAV
CQFVTNVNMSNGRVPLPTNRPNGTACPPSTSTSTPPSQQTIVVVENPMSVDESGLVSNV
VVGVTDDKK*

>G1512 (1..732)

ATGGAAGGGAACCTTCTTCATCAGGTCTGATGCTCAACGAGCACATGACAATGGCTTCATA
GCCAAACAAAAACCTAATCTCACCACGGCTCCAACAGCAGGTCAAGCTAATGAAAGTGGC
TGTTTTGACTGCAACATCTGTTTAGACACAGCCCATGATCCGGTGGTCACTCTCTGCGGG
CACCTTTTCTGCTGGCCTTGCAATTTACAAGTGGTTACATGTTTCACTTCTCTGCTCTCC
GTTGATCAGCACCAGAACAATGCCCTGTTTGTAAATCCAACATTACTATCACCTCTTTG
GTTCTCTCTATGGAAGAGGCATGTCTTCGCCTTCTTCCACGTTTGGCTCCAAGAAACAA
GACGCACTGTCCACTGACATACCCCGCAGACCTGCTCCATCAGCCTTACGCAATCCGATT
ACCTCAGCATCATCTCTGAACCCAAAGCTTGCAACATCAAACCTCTGTCTCTTTCATTTTCA
AATCATCAGTATTCCCCTCGTGGCTTCACCACAACCGAATCAACCGACCTTGCCAATGCT
GTAATGATGAGTTTCTCTACCTGTGATTGGAATGTTTGGAGACCTGGTCTACACCAGG
ATATTCGGGACCTTCACAAACACAATAGCTCAGCCTTACCAAAGCCAGAGGATGATGCAG
CGTGAGAAAGTCTCTTAATCGGGTATCGATATTCTTCTTTGTTGCATCATCCTTTGCCTC
CTTCTCTTCTAG

>G1512 Amino Acid Sequence (domain in AA coordinates: 39-93)

MEGNFFIRSDAQRAHDNGFIAKQKPNLTAPTAGQANESGCFDCNICLDTAHD PVVTLCG
HLFCWPCIIYKWLHVQLSSVSVDQHQNCPVKSNITITSLVPLYGRGMSSPSSTFGSKKQ
DALSTDIPRRPAPSALRNPIITSASSLNPSLQHQTLSPSFHNHQYSRPGFTTTESTDLANA
VMMSFLYPVIGMFGDLVYTRIFGTFTNTIAQPYQSQRMMQREKSLNRVSIFFLCIIILCL
LLF*

>G2069 (1..1026)

ATGGAAGGAGGAGGAAGAGGACCAAATCAAACGATTCTCAGTGAAATAGAACATATGCCT
GAAGCTCCACGTCAACGTATCTCTCATCACCGTCGAGCTCGCTCTGAAACCTTCTTCTCC
GGCGAATCAATCGACGATCTCTCTTATTCGATCCTTCCGATATCGATTTCTCTTCTCTA
GACTTCTCAACGCTCCACCACCACCAACAATCACAACAACACCGCAAGCTTCTCCC
ATGTCCGTTGATTCCGAAGAAACCTCATCGAACGGTGTGTTCTCTCTAATTCTCTTCTCT
CCAAAACCCGAAGCTAGATTCCGTCGCCATGTTCTGATGCTTCTCGGTTGATTCCGATTTC
TTCGATGATTTGGGTGTTACTGAGGAGAAGTTTATAGCTACAAGTTCAGGAGAGAAGAAG
AAAGGGAATCATCATCATAGCAGGAGTAATTTCTATGGATGGAGAGATGAGTTTCGGCGTCG
TTTAATATCGAATCGATTTTAGCTTCTGTGAGTGGTAAAGATAGTGGGAAGAAGAATATG
GGTATGGGTGGTGATAGACTTGCTGAGCTTGCTTTGCTTGATCCTAAAAGAGCTAAAAGG
ATTTTAGCGAATAGACAATCTGCGGCGAGGTTCGAAAGAGAGGAAGATTAGGTATACTGGT
GAGTTAGAGAGGAAGGTTTCAGACACTTCAGAATGAAGCTACTACATTGTCTGCTCAAGTC
ACTATGTTACAGAGAGGAACATCAGAGCTGAACACTGAAAATAAACACCTCAAAATGCGG
CTTCAAGCTTTAGAGCAACAAGCTGAACCTTAGGGATGCTTTGAATGAAGCGCTGCGGGAT
GAACTGAACCGACTTAAGGTGGTAGCTGGAGAAATTCCTCAGGGGAATGGAAATTTTAC
AACCGTGCTCAATTTCTCATCTCAGCAATCGGCAATGAATCAGTTTGGGAACAAAACGAAC
CAACAGATGAGTACAAACGGGCAGCCATCGCTCCCAAGCTACATGGATTTACCAAGAGA
GGCTGA

>G2069 Amino Acid Sequence (domain in AA coordinates: TBD)

MEGGGRGPNQITLSEIEHMPAPRQRISSHRRARSETFFSGESIDDLFLFDPSDIDFSSL
DFLNAPPPPPQQSQQQPQASPMVDSEETSSNGVVPNSLPPKPEARFGRHVRSFVSDFSDF
FDDLGVTEEFKFIATSSGEKKKGNHHHSRNSMDGEMSSASFNIESILASVSGKDSGKKNM
GMGGDRLAELALLDPKRAKRILANRQSAARSKERKIRYTGELBRKVQTLQNEATTLQAQV
TMLQRGTSELNTENKHLKMLRLQALEQQAE LRDALNEALRDELNRLKVVAGEIPQGNNGSY
NRAQFSSQQSAMNQFGNKTNQMSNTNGQPSLPSYMDFTKRG*

>G1852 (55..1857)

CATCTGATCTGCTCTCGAAGACGAAAGCTTCGAGTACTGGTTGAAGCTAAAGCTATGGGA
CACGTGAATCTACCTGCATCAAAGCGTGGTAACCCCTCGTCAATGGCGTCTCCTCGACATC
GTAACCGCTGCTTTCTTCGGTATCGTACTTCTCTTCTTCATCCTTTTATTCACTCCTCTT
GGTGATTCCATGGCGGCTTCTGGTCGGCAAACGCTGCTTCTCTCTACGGCGTCAGATCCG
AGGCAACGGCAGCGATTAGTGACTTTGGTTGAAGCTGGTCAGCATTTGCAACCGATCGAG
TATTGTCTCTGCGGAAGCTGTTGCTCATATGCCCTTGTGAGGATCCGAGAAGGAATAGTCAG
CTTAGTAGAGAGATGAATTTCTATAGGGAGAGACATTGTCTTTGCTGAGGAGACTCCG
CTCTGTTTGATTCTCTCCGCTTCTGGTTATAAAATTCCTGTTCCGTGGCCTGAGAGTCTT
CACAAGATTGGCATGCAACATGCCATATAACAAAATTGCTGACCGGAAAGGTCATCAA
GGATGGATGAAAAGGGAAGGGAATACTTTACTTTCCAGGCGGTGGCAGCATGTTTCTCT
GGCGGAGCTGGCCAATACATTGAAAAGCTTGACACAGTATATTCCGCTTAATGGTGGAACT
TTGAGAACTGCTCTTGACATGGGATGCGGGGTAGCTAGTTTGGAGGTACTCTACTATCT
CAAGGCATTCTAGCCCTCTCATTTGCTCCAAGAGATTACATAAATCTCAAATTCAGTTC
GCTTTGGAAAGAGGAGTGCCTGCATTGTTGTCATGCTTGGCACTCGTAGACTCCCTTTT
CCTGCATACTCTTTGACCTGATGCACTGTTCCCGATGTTTGATTCTTTTACGGCTTAC
AATGCAACTTACTTCATCGAAGTAGATAGGTTACTGCGCCCTGGAGGATATCTTGTAAATC
TCTGGCCCACCTGTACAATGGCCTAAACAAGACAAAGAATGGGCTGATCTTCAGGCGGTG
GCTAGAGCTTTGTGCTATGAGCTAATTGCGGTTGATGGAAACACTGTCATCTGGAAGAAG
CCTGTTGGAGATTCTAGTCTACCTAGCCAGAATGAGTTTGGGCTTGAGTTGTGTGATGAG
TCTGTTCCGCCAAGTGATGCATGGTATTTTAAATTGAAGAGGTGTGTTACCAGGCCATCA
TCCGTCAAAGGAGAACACGCTTTGGGAACTATATCCAAGTGGCCGGAGAGGCTTACTAAA
GTTCTCTTCTAGGGCATTGTGATGAAAACGGATTGGATGTGTTGAAGCAGATGCAAGG
CGGTGGGCAAGACCGCTTGCTTATTACAGGGATTCTCTTAACCTGAAGCTGAAATCTCCA
ACTGTCCGCAATGTCATGGACATGAACGCATTCTTCGGAGGCTTTCAGCAACCCCTTGCA
TCTGATCCTGTGTGGGTTATGAATGTCATTCCAGCTCGGAAGCCATTAACTCTTGACGTG
ATTTATGACAGAGGTCTCATCGGTGTTTACCATGATTGGTGTGAACCATTTTCAACATAT
CCCCGCACGTATGATTTTCATCCATGTATCAGGAATTGAATCAGTATAAACGACAAGAC
TCAAGCAAATCGAGGTGTAGCCTAGTAGATCTAATGGTAGAGATGGACAGAATATTACGT
CCAGAAGGAAAGGTTGTGATCCGAGACTCTCCTGAGGTGCTAGATAAAGTCGCACGAATG
GCTCATGCTGTAAGATGGTCTTCTTCCATACACGAGAAAGAACCTGAATCCCATGGAAGA

GAGAAGATTCTTATCGCAACCAAATCTCTCTGGAAATTGCCATCAAACCTCCCACTGAAGA
CACAAAAAGAAGAAGAAAAGAAGCTCTTCTCAATCTTGTAGGTACTGTCACTTGCTCT
CCAGCCC.

>G1852 Amino Acid Sequence (domain in AA coordinates: 1-601)
MGHVNLPASKRGNPRQWRLLDIVTAAFFGIVLLFFILLFTPLGDSMAASGRQTLLLLSTAS
DPRQRQRLVLTLEAGQHLQPIEYCPAEVAHMPCEDPRRNSQLSREMNFYRERHCPLPEE
TPLCLIPPPSGYKIPVPWPESLHKIWHANMPYNKIADRGKHQGWMKREGEYFTFPGGGT
FPGGAGQYIEKLAQYIPLNGGTLRTALDMGCGVASFGGTLLSQGILALSFAPRDSHKSQI
QFALERGVPAPFVAMLGTRRLPFPAYSFDLMHCSRCLIPFTAYNATYFIEVDRLLRPGGYL
VISGPPVQWPQDKQWADLQAVARALCYELIAVDGNTVIWKKPVGDSCLPSQNEFGLELC
DESVPSPDAWYFKLKRCVTRPSSVKGEHALGTISKWPERLTKVPSRAIVMKNGLDVFEAD
ARRWARRVAYYRDSLNLKLKSPFVRNVMDMNAFFGGFAATLASDPVWVMNVI PARKPLTL
DVIYDRGLIGVYHDWCEPFSTYPRTYDFIHVSGIESLIKRDSSKSRCSLVDLMVEMDRI
LRPEGKVVIDRSPVELDKVARMAHAVRWSSSIHEKEPESHGREKILITKSLWKLPNSNH

*

>G1793 (59..1783)

AGTGATTTATTGATTAACCCAAACACAAAATAAACAGATTTGACTCAAAAAGAAGAAAAT
GAATTCTAACAACCTGGCTTGGCTTTCTCTTTACCCGAACAACCTCTTCTTTCCTCTCA
TGAATACAACCTTGGCTTGGTCAGCGACCATATGGACAACCTTTTCAAACACAAGAGTG
GAATATGATCAATCCACACGGTGGAGGAGGAGATGAAGGAGGAGAGGTTCCAAAAGTGGC
CGATTTTCTCGGTGTGAGCAAAACCGGACGAAAACCAATCCAACCACCTAGTAGCTTACAA
CGACTCAGACTACTACTTCCATACCAATAGCTTGTAGCTAGCGTCCAATCAAACGATGT
CGTTGTAGCAGCTTGTGACTCCAATACTCCTAACACAGTAGCTATCATGAGCTTCAAGA
GAGTGCTCACAATCTACAGTCACTTACTTTGTCCATGGGGACCACCGCTGGTAATAATGT
TGTAGACAAAGCTTACCATCCGAGACCACCGGGGATAACGCTAGCGGTGGAGCACTAGC
CGTTGTTGAGACGGCCACGCCAAGACGTGCATTGGACACTTTCGGACAACGAACCTCGAT
CTATCGTGGTGTCAAGACATCGATGGACTGGTCGATATGAGGCTCATCTATGGGATAA
TAGTTGTAGAAGGGAAGGCCAGTCTAGGAAAGGAAGACAAGTTTACTTGGGTGGATATGA
CAAAGAAGATAAAGCAGCAAGATCATATGATCTAGCTGCACTTAAGTACTGGGGTCCTTC
AACTACTACTAATTTCCCAATTAACAACTACGAGAAAAGAGTAGAGGAAATGAAGCACAT
GACGAGACAAGAGTTCGTGGCTGCCATTAGAAGGAAAAGTAGTGGATTTTCGAGAGGCGC
TTCGATGTATCGAGGAGTTACAAGGCATACCAACATGGAAGATGGCAAGCAAGGATCGG
CCGAGTCGCGGAAACAAAGACCTCTACTTGGGAACCTTTAGCACTGAGGAAGAAGCAGC
AGAAGCTTACGATATAGCTGCAATAAAGTTTAGAGGACTTAATGCAGTGACCAACTTCGA
GATCAACCCGTACGACGTGAAAGCCATTCTAGAGAGTAGCACTCTTCCCATCGGAGGAGG
CGCAGCTAAACGGCTCAAAGAAGCTCAAGCTCTTGTAGTCTTCAAGGAAACGCGAGGCGGA
GATGATAGCCCTTGGTTCAAGTTTCCAGTACGGTGGTGGCTCGAGCACAGGCTCTGGCTC
CACCTCATCAAGACTTCAAGCTTCAACCTTACCCTCTAAGCATTCAACAACCATTAGAGCC
TTTTCTATCTCTTCAAGACAATGACATCTCTCATTACAACAACAACATGCTCACGATTC
CTCTCTTTTAAATCACCATAGCTATATCCAGACACAACCTTCATCTCCACCAACAGACCAA
CAATTACTTGCAGCAACAGTCGAGCCAGAACTCTCAGCAGCTCTACAATGCGTATCTTCA
TAGCAATCCGGCTCTGCTTCATGGACTTGTCTCTACCTCTATCGTTGACAACAATAATAA
CAATGGAGGCTCTAGTGGGAGCTACAACACTGCAGCATTTCTTGGGAACACGGTATTGG
TATTGGGTCCAGCTCGACTGTTGGATCGACCGAGGAGTTTCCAACCGTTAAAACAGATTA
CGATATGCCTTCCAGTGTGGAACCGGAGGGTATAGTGGTTGGACCACTGAGTCTGTTCA
GGGCTCAAACCCCTGGTGGTGTCTTCACTATGTGGAATGAGTAAACAAGGATCTCTTTCTT
GCGGCACAAGGAATGGGT

>G1793 Amino Acid Sequence (conserved domain in AA coordinates:179-255, 281-349)

MNSNNWLGFPLSPNNSSLPHEYNLGLVSDHMDNPFQTQEWNMINPHGGGDEGGEVVKV
ADFLGVSKPDENQSNHLVAYNDSDYFHTNSLMPSVQSNVVAACDSNTPNNSSYHELQ
ESAHLQSLTLMSGTTAGNNVVDKASPSETTGDNASGGALAVVETATPRRALDTFGQRTS
IYRGVTRHRWTGRYEAHLWDNSCRREGQSRKGRQVYLGYYDKEDKAARSYDLAALKYWGP
STTTNFPIITNYEKEVEEMKHMTRQEFVAAIRKSSGFSRGASMYRGVTRHHQHRWQARI
GRVAGNKDLYLGTFTSTEEAAEAYDIAAIKFRGLNAVTFEINRYDVKAILESSTLPIGG
GAAKRLKEAQALESSRKREAEMIALGSSFQYGGGSSSTGSGSTSSRLQLQPYPLSIQPPLE
PFLSLQNNDISHYNNNNNAHSSSFNHHSYIQTQLHLHQQTNNYLOQQSSQNSQQLYNAYL
HSNPALLHGLVSTSIVDNNNNNGSSSGSYNTAAFLGNHGIGIGSSSTVGSTEEFPTVKTD

YDMPSSDGTGGYSGWTSESVQGSNPGGVFTMWNE*

>G761 (521..1549)

GGGGCCGACCGGCCCGGGGCGAGGTCTAGGTTCAAAAGGACTCACAAGAGAGAGATAGT
ATGATTGATAGGGAAAGAGAGAGAGATGAAAGAAAGTAAAATATATAATAGATTATTAGG
ACACGAGTGTCTATCTTTTGATTTGTGTCTTGTGTGCTCTCTTTCTCTCTTCTCTCGAA
TGATCATCTTTATATAACCCCTACTCTCTTTCTCTTTTCCCATTCTTTTCATATCATTTCTCC
CTTTCTCTCTCGGGATCTGATCTCTCTTTCCAGTAACCTATTCCCGAGGAGCACTGTCAA
ATCTTGTCCACTCTTTGATCTTATCTCGATCTCTTTCTCTTTCTAGTCTTGTGTAGTCTT
CAAAGTGTGATGTTATCTATATAGTAATCACGAGAGAGAATCATACAATAGCTGAAACA
TAAAGCTTTCTTAGAAGCTTTAAAAAGGTCTCATCTGGATTATCCTGTTTAAATTTCTAGA
GTTTCTTCAGGCAGATTATTAACCGATCAAGAAGACAAACATGAATTCATTTTCCACGT
CCCTCCGGGTTTTAGATTTACCCGACAGATGAAGAACTTGTAGACTACTACCTGAGGAA
AAAAGTCGATCGAAGAGAATAGAAATTGATTTTATAAGGACATTGATCTTTACAAGAT
TGAGCCATGGGACCTTCAAGAGTTGTGCAAAATTGGGCATGAAGAGCAGAGTGATTGGTA
CTTCTTTAGCCATAAAGACAAGAAGTATCCACAGGGACTCGAACCAATAGAGCAACAAA
AGCAGGGTTTTGAAAGCCACCGGAAGAGATAAGGCTATCTATTTGAGGCATAGTCTAAT
TGGCATGAGGAAAACACTTGTGTTTTACAAGGAAGAGCCCCAAATGGACAAAAGTCTGA
TTGGATCATGCACGAATACCGCTTAGAAAACCGATGAAAACGGAACTCCTCAGGAAGAAGG
ATGGGTTGTGTGTAGGGTTTTCAAGAAGAGATTGGCTGCAGTTAGACGAATGGGAGATTA
CGACTCATCCCTTCACATTGGTACGATGATCAACTTTCTTTTATGGCTCCGAGCTCGA
GACAAACGGTCAACGACGGATTCTCCCAATCATCATCAGCAGCAGCAGCAGCAGCAGCA
ACAACATATGCCATATGGCCTCAATGCATCTGCTTACGCTCTCAACAACCTTAAGTGA
ATGCAAGCAAGAGCTAGAATACTACACTACAACCACCTGCAATCAAATATCGCGCATGAGGA
ACAATTGAATCAAGGAAATCAGAACTTCAGCTCTCTATACATGAACAGCGGCAACGAGCA
AGTGATGGACCAAGTCACAGACTGGAGAGTTCTCGATAAATTTGTTGCTTCTCAGCTAAG
CAACGAGGAGGCTGCCACAGCTTCTGCATCTATACAGAATAATGCCAAGGACACAAGCAA
TGCTGAGTACCAAGTTGATGAAGAAAAAGATCCGAAAAGGGCTTCAGACATGGGAGAAGA
ATATACTGCTTCTACTTCTTCGAGTTGTCAGATTGATCTATGGAAGTGAGCTGAAAGAGA
AGACATATAAATGCATATATACATATATATATATACGTACACAGAACAATAATCAAGTG
TAGATGATGATGATGGTACAGATTTATATTTGCTTTGATTGATTCTTACTACATTATTGA
ACTTATGTCATATGCATATATACATTGCGTATCTATGCATATTTATACTTGTACTCAATA
TGATTAACCATATATAAACTCTAATCTAAATGTAACCTCAATATTTTTTAAATAGACAAT
TGTCTCTTCTTATTAGAAAAAAA

>G761 Amino Acid Sequence (domain in AA coordinates: 10-156)
MNSFSHVPPGFRFHPTDEELVDYYLRKKVASKRIEIDFIKDIDLYKIEPWDLQELCKIGH
EEQSDWYFFSHKDKKYPTGTRTNRATKAGFWKATGRDKAIYLRHSLIGMRKTLVFKGRA
PNGQKSDWIMHEYRLETDENGTPQEEGWVVRVFKRLAAVRRMGDYDSSPSHWYDDQLS
FMASELENTGQRRILPNHHQQQHEHQHMPYGLNASAYALNNPNLQCKQLELHYNHLQ
SNIAHEEQNLQNQNFSSLYMNSGNEQVMDQVTDWRVLDKFVASQLSNEEAATASASIQN
NAKDTNSAEYQVDEEKDPKRASDMGEBYTASTSSSCQIDLWK*

>G1056 (10..798)

GCTACATATATGGGTTCTATTAGAGGAAACATTGAAGAGCCTATATCTCAGTCATTAACG
AGGCAGAACTCTCTCTATAGCTTAAAGCTCCATGAGGTTCAAACCCACTTAGGAAGTTCT
GGAAAACCACTAGGAAGCATGAACCTTGATGAGCTTCTCAAGACTGTCTTGCCACCAGCT
GAGGAAGGGCTTGTTCGTGAGGAAGCTTGACGTTACCTCGAGATCTCAGTAAAAAGACA
GTTGATGAGGCTCGGAGAGATATCCAACAGGACAAGAATGGAACGGTACTAGTACTACT
ACTACTCATAAGCAGCCTACACTCGGTGAAATAACACTTGAGGATTTGTTGTTGAGAGCT
GGTGTAGTGACTGAGACAGTAGTCCCTCAAGAAAATGTTGTTAACATAGCTTCAAATGGG
CAATGGGTTGAGTATCATCATCAGCCTCAACAACAACAAGGGTTTATGACATATCCGGTT
TGCGAGATGCAAGATATGGTGTATGATGGGTGATTATCGGATACACCACAAGCGCCTGGG
AGGAAAAGAGTAGCTGGAGAGATTGTGGAGAAGACTGTTGAGAGGAGACAGAAGAGGATG
ATCAAGAACAGAGAATCTGCAGCACGTTACGAGCTAGGAAACAGGCTTATACACATGAA
TTAGAGATCAAGGTTTCAAGGTTAGAAGAAGAAAACGAAAACCTTCGGAGGCTAAAGGAG
GTGGAGAAGATCCTACCAAGTGAACCACCACAGATCCTAAGTGGAAGCTCCGGCGAACA
AACTCTGCTTCTCTCTGATCCTAAAGACTCTTCTTTCTTTCTTCTTCTTGTGTTGGTTT
ATATCAGACCGCTTTGTTCTTTGTATATTGTGTAGACTTTATGACTTTGAACAGCATGT
CTTTATAAACATTTCTTGAGTGT

>G1056 Amino Acid Sequence (domain in AA coordinates 183-246)
MGSIRGNIEEPISQSLTRQNSLYSLKLHEVQTHLGSSGKPLGSMNLDELLKTVLPPEEG
LVRQGSLLTPRDLSSKKTVDDEVWRDIQQDKNGNGTSTTTTHKQPTLGEITLEDLLLRAGVV
TETVVPQENVVNIASNGQWVEYHHQPQQQGFMTYPVCEMQDMVMMGGLSDTPQAPGRKR
VAGEIVEKTVERRQKRMIKNRESAARSARKQAYTHELEIKVSRLEENELRRLKEVEK
ILPSEPPDPKWKLRRTNSASL*

>G1447 (82..1086)
AAAAACCTAACCTAATTCTCTCAAGACAACTCAAAGGTCTCTCCTTTTTTAGGTTTAT
TATCACTTCCGTATAATCGCCATGTCTTCTCTACCATGGAAAAACCAAATCGAGTCGA
ATCTTAAGATTCAATTTCTGAGTTTCAACAATCACCGTTCGTTGAAACTGGCTTTTCCAACT
TCTCTGATCGATCTCTTCTTCAAGAATCGCGATCGTCTAAAAAATCTCCATCTAAACGC
TTCCAACGAATCGAACGCCAGATTCTGAACCGCTCCAAACGCTTCTTCGTTGAGTAATCAA
GATACGATTTTTGAAAAGCCCTCGAGGATTAAACCGTTCGAAGTAAGGTGAGAAAAGTT
AATTGCGTTAAAGGTAAATCAGCGCGTGAAGAAGAACGCGATTAAAAATAGCGTTTTTC
GGCGGTAGCGGTGAGGTCTTTTGTGCGCTTTAAGGTTTTGATAGTAGCGTTGCTCGCC
TTGAGCACGAAGAAGAAGCTCACTTTAGGAATCACTCTCTCTGCTTCGCTCTTCTCTTA
ACAGAGCTCGTGGCGGCGGTGTTTTACGCGCTCTAATAACACCGACAAAGACAAAAAC
GCGATTGCCCCGAGAAAAATCGAAACTTTTGATGAAACTCGAGTTCCCAAAGCGATTCCA
TGTCCTGAGGAAACAGAGCATGTAGTATCTGAAACAGAGGTTTCGAAGTTGAAAGGTTTA
ACGATACGTGATCTGTTGTCAAAGGACGAGAAATCAACAAGTAAAGTTGGAGACTAAAA
TCGAAGATTGTGAAGAAGTTGAGGAGTTACAATAAGAAGGATAAGAAGACGATGAAGATC
AAAGAAGAGTCTTTGATTGAAGTCTCGAGTTTGGTTTATAGAAGATAAACCAAAGAAAATT
GAGTCTGAGAGAGACGAAGAAGAAACGTTGAATCCTCCAGTGGTTGGATCAAACCTGAAT
GGGATTGTTCTGATCGTGATTGTGCTAACCAGTTTGTATGTGGGAAGGTCTTAGCTATT
GTTCTGACACTATCATGTTTGGTTCTTAGATTAGGAGCAGTCAAAAAAGTTAATCTTTGC
ATATAATTTTTTTTGTATTTTAAACATGCTTGCATGTGAAACTGTAAATTTTTCTCATT
CATATGAAGGAGATTGGATTGAATGTTGAATACTAAA

>G1447 Amino Acid Sequence (domain in AA coordinates: 3-54, 124-156)
MSSLPWKKPKSSRLRFISEFQQSPFVETGFPTSLIDLFFKNRDLKKSPSKRFQRIERQ
IRTAPNASSLSNQDTIFKEKPSRIKTVRSKVEKVNCKVKSAAALKKNAIKNSVFGGSSEVV
LMAFKVLIVALLALSTKKKLTGILTSALFALLTELVAARVFRSNNTDKDKNAIAREKI
ETFDETRVPKAIKPCPEETHEVSVSETEVSKLGLTIRDLKSKDEKSTSKSWRLKSKIVKKL
RSYNKKDKKTMKIKESLIEVSSLVLEDKPKKIESERDEBETLNPVVGSNLNGIVLIVI
VLTGLLCGKVLAIVLTLSCVLRLGAVKKVNLCI*

>G323 (77..826)
CTGCTCATATCAGCCATTGACACAGTTGCTTTGGGTTTCCCTCAAACGGCGCCGATTGTC
TGGATTTTGACCACTGATGGCCTTAGATCAATCTTTGAAGATGCTGCTTTACTTGGAGA
ACTCTATGAGCAAGGTGCATTTTGTTCAGAGCAAGAACTGAACCCATTACAGTCTC
GGTTCCTTCTGATGATACTGATGATTGCAATTTTGAAGTCAATATTTGCTTAGACTCGGT
GCAAGAACCTGTTGTGACTCTCTGTGGTCACCTCTTTTGTGCTGGCCTTGTATTCACAAATG
GCTTGATGTACAGAGCTTCTCAACAAGTGATGAATACCAAGACATAGACAGTGTCTGT
TTGTAAATCTAAAGTTTCTCATTCTACTTTGGTTCCCTTGTATGGTAGAGGCCGTTGTAC
TACTCAGGAGGAAGGTAAAAACAGTGTGCCTAAAAGACCCGTAGGACCGGTTTATCGGCT
TGAAATGCCGAATTCACCTTATGCAAGTACTGATCTGCGGTTATCACAACGGGTTTCAATTT
CAATAGCCACAGGAAGTTACTACCTGTCTCAGGGGTGATGAGCTCGAACAGTTTATC
ATACTCTGCTGTTTTTGGATCCGGTGATGGTGATGGTTGGAGAAATGGTAGCTACGAGGTT
GTTTGGAAACACGAGTGATGGATAGATTGCGTATCCGGACACTTACAATCTCGCAGGGAC
TAGCGGGCCGAGGATGAGAAGGCGGATAATGCAGGCAGATAAATCGCTGGGAAGAATCTT
CTTCTTCTTTATGTGTTGTGTGTTCTGTGCTTCTCTGTTTTAGGTTTTCATAGCTAG
CTTGGTTCTGCTACTGTTCACTTCTTCAGG

>G323 Amino Acid Sequence (conserved domain in AA coordinates:48-96)
MALDQSFEDAALLGELYGEGAFCKSKKPEPITVSVPSDDTDDSNFDCNICLDSVQEPVV
TLCSGHLFCWPCIHKWLDPVQSFSTSDYQRRHQCPVCKSKVSHSTLVLPLYGRGRCTTQEEG
KNSGPKRPVGPVYRLEMPNSPYASTDLRLSQRVHFNSPQEGYYPVSGVMSNSLSYSAYL
DPVMVMVGEMVATRLFGTRVMDRFAYPDTYNLAGTSGPRMRRRIMQADKSLGRIFFFMC
CVVLCLLLF*

>G176 (41..1606)

AGAAGAAGAAGAAGAGTACCTCATACGTAACCATTGATGGGCTCTTTTGATCGCCA
AAGAGCTGTTCCGAAATTCAAAAAGCAACACCGTCACCGCTCCCTCTTCTCCTTCGCC
TTACTTCACTATGCCTCCTGGCCTTACTCCCGCCGACTTCTCGACTCTCCTCTTCTCTT
CACTTCCTCCAACATTTTGGCGTCTCCTACGACAGGCACATTTCCAGCGCAATCTCTGAA
CTATAACAATAACGGTTTGCCTCATTGACAAAAATGAAATCAAATATGAAGACACAACCTCC
TCCCTTGTTCTTACCATCTATGGTAACTCAGCCTTACCTCAACTGGATTTATTCAAATC
CGAAATCATGTGAGTAACAAAACCTCTGATGACGGCTACAATTGGCGCAAATACGGGCA
GAAGCAAGTCAAAGGAAGCGAAAACCCGAGGAGTTACTTCAAATGCACGTATCCAAATTG
TCTCACAAGAAGAAAGTAGAGACGTCTTGTGAAGGGTCAGATGATTGAGATTGCTTA
TAAAGGAAGCCACAATCATCCCAGCCCCAATCCACGAAGCGATCATCTCCACCGCTAT
AGCAGCACATCAGAACAGCAGTAATGGAGACGGTAAAGACATTGGTGAAGATGAAACAGA
GGCCAAGAGATGGAAAAGAGAAGAGAATGTGAAGGAGCCAAGAGTGGTGGTTCAGACAAC
AAGTGATATAGACATTCTTGACGATGGCTACAGATGGAGAAAGTATGGTCAGAAAGTCGT
CAAGGGTAATCCAAATCCAAGGAGCTATTACAAGTGCACATTTACAGGATGTTTTGTAAG
GAAACACGTTGAAAGAGCATTCAAGATCCCAAGTCAGTGATCACAACCTACGAAGGAAA
ACACAAACACCAAATCCCGACCCCCAAGAAGAGGTCCAGTTTAAAGATCTGCTGCAATGGC
TTCTCCTCTTCTCCCAACTTCGACTACTCCTGATCAACTTCCCGGCGGCGATCCACAGTT
GCTGAGCTCTTACGCGTCTCTTGTCTCCCGGTTCTAGCCACCGTCCGTACGCTTCTGC
AGATGCCAGACCCTGGGCAGAGCTCGTTGACCGGTACGCGTTTTCCCGGCCACCATCGCT
CTCGGAGGCAACGTACGAGTAAGGAAGAACTTTTCTATTTCCGAGCCAATTACATAAC
CTTAGTGGCAATCTTACTCGCCGCTCTCTGCTCACGACCCCTTTCGCTCTCTTCTCCT
CGCATCGCTGGCCGCTTCTTGGCTTTTCTCTACTTTTTCGCTCCGCGGATCAGCCGTT
GGTCATTGGAGGACGCACGTTCTCCGATCTTGAGACGCTAGGGATACTCTGCCTGTCCAC
TGTGGTGGTGATGTTTCATGACCAGCGTTGGATCGCTCTTGATGTCCACTCTAGCAGTTGG
GATCATGGGCGTGGCCATCCACGGAGCGTTTCGTGCTCCCGAAGACCTGTTCTTGAAGA
ACAAGAAGCCATTGGATCTGGACTTTTCGCATTCTTCAACAACAATGCCCTAATGCAGC
TGCCGCTGCCATAGCCACCTCAGCAATGTACGCGTTTCGAGTCTGAGATTGTTGAAGAGA
CTACATTCCTACACCGCATTTCCAAAGTGTGATATTTATTCATATTGAATTGTT
>G176 Amino Acid Sequence (domain in AA coordinates: 117-173, 234-290)
MGSFDRQRAVPKFKTATPSPLPLSPYFTMPPLTPADFLDSPLLFTSSNILPSPTTGT
FPAQSLNYYNNGLLIDKNEIKYEDTTPPLFLPSMVTQPLPQLDLFKSEIMSSNKTSDG
NWRKYQKQVKGSENPRSYFKCTYPNCLTKKKVETSLVKGQMIIEIVYKGSNHPKPQSTK
RSSSTAIAAHQNSSNGDGKDIDEGDETEAKRWKREENVKEPRVVVQTTSDIDILDDGYRWR
KYGQKVVKGNNPNRSYYKCTFTGCFVRKHVERAFQDPKSVITTYEGKHKHQIPTPRRG
LRSAMASPLLPSTTPDQLPGGDPQLSSLRVLLSRVLATVRHASADARPWAEVLVDRSA
FSRPPSLSEATSRVRKNFSYFRANYITLVAILLAASLLTHPFALFLLASLAASWLFYFF
RPADQPLVIGRFTFSDLETGLILCLSTVVMFMTSVGSLMSTLAVGIMGVAIHGAFRAP
EDLFLLEEQAETGSLFAFFNNNASNAAAAIAATSAMSRVRV*
>G174 (194..1585)
CCCAATTTGAGATTGTTTCGATTTTCGATCTACGAGATTCTTACAAGAACATAAGCAGCTTC
GGTTTTTTTGGGATTATCTTATTTGGTCGGATGATGATCTTCTCGATGTCTGTGCTAGGCT
TTGGGAATTAGATATATTTGGGGTTAAGCTCGAGTCTCTCCGGTTTTTGAGTTTACTTGAG
TTTGTTAGTATTTATGGCTGAGGTGGGAAAAGTTCTGGCTAGTGATATGGAGTTAGACCA
TTCAAATGAGACTAAAGCAGTGGATGATGTTGTTGCCACTACTGATAAAGCGGAGGTCAT
ACAGTGGCTGTAAGTAACTAGAACTGAAACCGTTGTTGAAAGTTTGGAATCTACTGACTGTAA
GGAGCTTGAAAAACTTGTTCCACATACGGTAGCTTCGCAGTCGGAAGTAGATGTTGCTTC
CCCGGTATCCGAGAAAAGCACCAGGTTTCTGAAAGTAGCGGTGCATTATCTTTGCAGTC
TGGTTCGGAAGGGAATAGTCTTTTTATTCTGTGAGAAGGTTATGGAAGACGGATACAACTG
GCGGAAATATGGACAGAACTTGTGAAAGGAAATGAGTTTGTAAAGGAGCTATTACAGGTG
CACTCACCTTAAGTCAAAGCGAAAAACAGTTGGAACCGTCTGCGGGTGGACAGTCGT
GGATACCGTTTTACTTTGGGGAACATGATCACCCAAAGCCTCTTGCTGGTGTGTTCTCTAT
CAATCAGGATAAGCGAAGTGATGCTTTCACAGCTGTTAGTAAAGAGAAAAACATCTGGATC
CAGTGTTCAGACACTTCGTCAAACCGAACCAACCAAGATCCATGGAGGATTACATGTTTC
AGTTATTCCACCACTGATGATGTGAAAACCTGATATTTCAATCAAGTAGGATAACGGG
GGACAACACTCACAAGGATTATAATAGTCCTACCGCAAAGCGAAGGAAGAAAGGAGGGAA
CATTGAGCTGAGTCCAGTGGAGAGGTCAACCAATGATTACGCATTGTGGTTCACACTCA
GACTCTGTTTGATATGTGAATGATGGGTACCGATGGCGTAAATATGGTCAGAAATCAGT

AAAAGGCAGCCCATATCCAAGGAGCTACTATAGATGTTCAAGCCCTGGATGCCCCGTCAA
GAAACACGTAGAGAGGTTCATCTCATGACACAAAGTTGCTTATAACAACCTTACGAGGGAAA
ACACGACCACGATATGCCTCCAGGAAGAGTTGTTACTCATAATAACATGCTGGACTCGGA
AGTTGATGATAAAGAAGGAGATGCCAACAAGACTCCACAGAGCTCAACTCTTCAATCCAT
TACAAAGACACGAGCATGTTCGAAGATCACTTAAGAAAGAAAACGAAGACTAATGGCTTTTGA
GAAAAGTCTTGTATCAAGGTCCAGTTTGGATGAGAAGCTGAAGGAGGAAATAAAAGAGAG
ATCAGATGCAAAACAAGATCACGCAGCCAATCACGCCAAGCCGGAAGCAAAGTCAGATGA
TAAACCCTGTTTGTCAAGAGAAGGCAGTAGGAACCTGGAGAGCGAGGAACAAAAACC
CAAGACAGAGCCTGCCCAAAGCTAAGCATTTCAGTGTGTACCGAGTGGTAATTTATATGG
CTGTTTTAACATAGATTAGTACAGCGCATATGGTTATAGACTGTACAGTTGTTGTTTCAGG
CGGGACCAGATTTAGATTAGTGTTAATGGAATAGTATGCTTTAATACCTTTATGTAACC
ACTTCCATTTGGTTCAAATAAGAGTTACAGGAAGAGAAGGTAACACAACAAGAGCCCTTC
TTTGTGTAGAGCCTGTGTAAATAGTTGTAGCATGGGGATGTATATGATTGATTCAACC
TTATTAATGGTTATGAGACAAAACCTATC

>G174 Amino Acid Sequence (domain in AA coordinates: TBD)

MAEVGKVLASDMELDHSNETKAVDDVVATTDKAEVIPVAVTRTETTVESLESTDCKELEK
LVPHTVASQSEVDVASPVSEKAPKVSESSGALSLSQSGSEGNPFIREKVMEDGYNWRKYG
QKLVKGNEFVRSYYRCTHPNCKAKKQLERSAGGQVVDTVYFGEHDHPKPLAGAVPINQDK
RSDVFTAVSKEKTSVQTLRQTEPPKIHGGLHVSVIPADDVKTDISQSSRITGDNTH
KDYNSTAKRRKKGNIELSPVERSTNDSRIVVHTQTLFDIVNDGYRWRKYGQKSVKQSP
YPRSYRCSPPGCFVKKHVERSSHDTKLITTYEGKHDHDMPPGRVVTNNMLDSEVDDK
EGDANKTPQSSTLQSIKDKQHVEDHLRKKTKTNGFEKSLDQGPVLDEKLKEEIKERSDAN
KDHAAHNAKPEAKSDDKTTVCQEKAVGTLSEEQPKTEPAQS*

>G715 (1..705)

ATGGATACCAACAACCAGCAACCACCTCCCTCCGCCGCCGAATCCCTCCTCCACCACCT
GGAACCACCATCTCCGCCGAGGAGGAGGAGCTTCTTACCACCACCTTCTCCAACAACAA
CAACAACAGCTCCAACATTTCTGGACCTACCAACGCCAAGAGATCGAACAAGTTAACGAT
TTCAAAAACCATCAGCTTCCACTAGCTAGGATAAAAAAGATCATGAAAGCCGATGAAGAT
GTTTCGTATGATCTCCGAGAACGACCGATTTCTCTCGCAAAGCTTGTGAGCTTTTCATT
CTCGAGCTCAGCATCAGATCTTGGCTTTCACGCTGAGGAGAATAAACGTCGTACGCTTCAG
AAAAACGATATCGCTGCTGCGATTACTAGGACTGATATCTTCGATTTCTTGTGATATT
GTTCTTAGAGATGAGATTAAGGACGAAGCCGAGTCTTCGGTGGTGAATGGTGGTGGCT
CCTACCGCGAGCGCGTGCCTTACTATTATCCGCCGATGGGACAACCAGCTGGTCTCGGA
GGGATGATGATTGGGAGACCAGCTATGGATCCGAATGGTGTATGTCCAGCCTCCGTCT
CAGGCGTGGCAGAGTGTTTGGCAGACTTCGACGGGGACGGGAGATGATGTCTCTTATGGT
AGTGGTGAAGTTCCGGTCAAGGGAATCTCGACGGCCAAGGGTAA

>G715 Amino Acid Sequence (domain in AA coordinates: 60-132)

MDTNNQPPPSAAGIPPPPGTTISAAGGASVHLLQQQQQLQLFWTYQRQIEQVND
FKNHQLPLARIKKIMKADEDVRMISAEAPILFAKACELFILELTIRSWLHAEENKRRTLQ
KNDIAAAITRTDIFDLVDIVPRDEIKDEAAVLGGGMVVAPTASGVPPYPPMPGQAPGPG
GMMIGRPAMPDPNGVYVQPPSQAWQSVWQTSTGTGDDVSYSGSGSSGQGNLDGQG*

>G588 (196..1599)

ATCTGAAGTGAACCAAGCTCAGGTTTTGTCTTCTCTTTGATCATTCCTTTCTCAGCAATA
TAAATTAGAGTTATATCCTTTATAAAGGATTTTGCTTTTTACCAACAACCCCTAAATTC
GGTGTCTCAGCAAGAATCACGTGATTCTCGTTCCTCTTCTCAGAAACCCATCATCTTC
TATCTCATTGAGAAATGGGTCAAAGTTTTGGGAGAATCAAGAAGATCGAGCGATGGTT
GAATCCACCATAGGCTCTGAAGCTTGCAGCTTTTTCATCTCAACAGCTTCAGCTTCCAAC
ACTGCCTTGTCCAAAGCTTGTCTCACCACCAAGTGATTCCAATCTCCAACAAGGGTTACGT
CACGTTGTTGAAGGATCTGATTGGGATTATGCTCTTTTCTGGCTAGCGTCCAACGTTAAT
AGCTCTGATGGTTGTCTTGTATCTGGGGAGATGGTCATTGCCGTGTCAAAAAGGGTGCT
TCAGGTGAGGATTACTCTCAGCAAGATGAGATCAAAAGACGTGTGCTTCGCAAGCTTCAC
TTGTGCTTCTGTTGTTTCAGATGAAGATCATCGTTTGGTGAATCAGGAGCTCTTACTGAT
CTCGACATGTTTTATCTGGCTTCTTTGTACTTTTCTTTAGGTGTGATACCAATAAGTAC
GGTCTGCTGGAACCTATGTGTCTGGGAAGCCTCTTTGGGCTGCAGATTTGCCTAGCTGC
TTGAGTTATTATAGGGTTAGGTCTTTCTTAGCTAGGTCAGCTGGTTTTTCAGACTGTGTTG
TCTGTACCAGTGAATCTGGAGTTGTGGAGCTTGGTCTTTAAGACATATTCCAGAAGAT
AAGAGTGTGATTGAGATGGTGAAATCAGTGTGGTGGTCTGACTTTGTTTCAGGCTAAA

GAAGCTCCTAAAAATCTTTGGTTCGACAGCTGAGTCTTGGTGGAGCAAAACCTCGGTCTATG
AGTATTAATTTCTCCCCGAAGACCGAGGATGACACGGGTTTCTCATTGGAATCGTATGAG
GTGCAAGCGATCGGAGGCTCTAATCAAGTGTATGGTTATGAGCAAGGGAAGATGAGACA
TTGTATCTAACTGACGAGCAAAAGCCGAGGAAGAGAGGGAGAGAAAACAGCAAATGGAAGA
GAAGAGGCTCTAAACCATGTGGAAGCGGAACGGCAGAGGAGGGAGAAGCTGAACCAGAGA
TTCTACGCTTTGAGAGCGGTGGTGCCTAACATCTCCAAGATGGACAAGGCTTCGCTCCTT
GCAGACGCAATCACTTACATCACGGATATGCAGAAGAAAATCAGGGTGTATGAAACAGAG
AAGCAGATAATGAAGAGGAGGGAGAGTAATCAGATAACTCCAGCAGAGGTTGATTATCAA
CAGAGGCATGATGATGCAGTTGTAAGGCTAAGCTGTCCGTGGAAACTCATCCAGTTTCA
AAGGTGATACAAACGTTGAGGGAGAATGAAGTTATGCCTCATGATTCCAACGTGGCCATC
ACAGAGGAGGGTGTGGTTCACACATTCCTCTCCGGCCTCAGGGTGGCTGCACCGCTGAG
CAGTTGAAGGACAAGCTCCTTGCCTCTCTATCACAGTAATATCACAGCAGTAAGTGTCTA
TGTAATAAGTGTAAACCGTGTGGAGGTTGTATCAATGTACTATTGCAAGCCAACCAAAAA
AAACTCCAGCTTAGTAGGATCGTGAATTTTCTTATATGTAATGTTGAGATTGTCTTT
TACATATAAAGATTGA

>G588 Amino Acid Sequence (domain in AA coordinates: 309-376)

MGQKFWENQEDRAMVESTIGSEACDFFISTASASNTALSKLVSPPSDSNLQQGLRHVVEG
SDWDYALFWLASNVNNSDGCVLWGDGHCVRVKKGASGEDYSQQDEIKRRVLRKLHLSFVG
SDEHRLVKSGALTDLDMFYLASLYFSFRCDTNKYGPAGTYVSGKPLWAADLPSCLSYYR
VRSFLARSAGFQTVLSVPVNSGVVLEGLSLRHIPEDKSVIEMVKS VFGGSD FVQAKEAPKI
FGRQLSLGGAKPRSMSINFSPKTEDDTGFSLESYEVQAIGGSNQVYGYEQKDETLYLTD
EQKPRKRGRKPANGREELNHNHVEAERQRREKLNQRFYALRAVVPNISKMDKASLLADAIT
YITDMQKKIRVYETEKQIMKRRESNQITPAEVDYQQRHDDAVVRLSCPLETHPVSKVIQT
LRENEVMPHDSNVAITEEGVVHTFTLRPQGGCTAEQLKDKLLASLSQ*

>G1758 (69..677)

GTCCCTCCTCTTAGCTTCAACCGCCGGAAGAACTAAACAACCTTCTTGGAAAAAAGAGA
AACTAAAAATGAATATCCTTCAAAACCTAACCTAGCTCCACAGATTTCACTGAATTTT
TCAAGTTCGATGATTTTGACGATACTTTTGAGAAGATCATGGAAGAAATCGGCCGTGAGG
ACCACTCGTCGTCACCGACTTTGAGTTGGAGTTTCATCGGAAAAGTTAGTGGCTGCAGAAA
TCACAAGCCCGCTTCAAAACAAGCCTAGCTACCTACCTATGAGCTTTGAAATAGGTGACA
AAGATGAAATCAAAAAGAGGAAGAGACACAAAGAAGATCCGATTATTCACGTCTTCAAAA
CGAAATCATCAATTGATGAAAAGGTTGCTTTAGATGATGGGTATAAATGGAGGAAATACG
GAAAGAAGCCGATAACGGGTAGTCCATTTCCAAGGCATTATCACAAGTGTTCGAGCCCAG
ATTGCAACGTGAAGAAGAAGATCGAAAGAGATACGAACAATCCAGATTACATATTGACAA
CATACGAAGGTAGACATAACCACCCAGCCCTTCTGTAGTTTATTGTGATTACAGCGACT
TTGATCTTAACTCTCTCAACAATTGGTCCCTTTCAGACGGCAAATACGTATAGTTTCTCTC
ATTCTGCTCCATATTGATCGATCGTAGTTACAAGTTTGTGTATATAGATGTATATATATA
TATCACCAATTACCATCGTAATCACGTCTCACATGTAACCTACGTACATATATCTTGTTT
GGGGTTGCTTTTGTATGTATTGAATTGGTGGAGGTAGAATGGAAGTCATCTTGATAGT
TGTAAGTTGATGTAAGGTTTGTATAGTCATTTTTTATAAAGTAACTAATTTGTACAA

>G1758 Amino Acid Sequence (domain in AA coordinates: TBD)

MNYPSPNPNSSTDFTEFFKFDDFDDTFEKIMEEIGREDHSSSPTLSWSSESSEKLVAAEITS
PLQTSLATSPMSFEIGDKDEIKKRKRHKEDPIIHVFKTKSSIDEKVALDDGYKWRKYGKK
PITGSPFPFRHYHKCSSPDCNVKKIERDTNPNPDYILTTYEGRHNHPSPSVVCDSDDFDL
NSLNNWSFQTANTYSFSHSAPY*

>G2148 (66..737)

GTCTCTAATATAAGCTTGAACGTTGCTATATATAAATGTAAAGGCGAACGCATAAGAAAA
GAAAAATGGAGAATGAAGCTTTTGTAGATGGTGAATTGGAGTCTCTTTTGGGGATGTTCA
ACTTTGATCAATGTTTCATCTAACGAATCGAGCTTTTGCAATGCTCCAAATGAGACTGATG
TTTCTCTTCTGATGATTCTTCCCATTTGGTACAATTCTGCAAAGTAACTATGCGGCCG
TTCTTGATGGTTCCAACCAACCAACGAACCGAAATGTGCACTCAAGACAAGATCTGTTGA
AACCAAGGAAGAAGCAAAAGTTAAGCTCGGAAAGCAATTTGGTTACCGAGCCTAAGACTG
CTTGAGAGATGGTCAAAGCCTAAGCAGTTATAATAGTTTCAAGATGATGAAAAGGCTTTAG
GTTTAGTGCTAATACATCAAAAAGCCTAAAACGCAAAAGCGAAAGCCAACAGAGGGATAG
CTCCGATCCTCAGAGCCTATACGCTAGGAAAACGAAGAGAAAGGATAAACGATAGGCTAA
AGACATTGCAGAGCCTAGTTCCTAATGGGACAAAGGTCGATATAAGCAATGCTGGAAG
ATGCTGTCCATTACGTGAAGTTCCTGCAGCTTCAAATCAAGCTCTTGAGTTTCAAGATC

TATGGATGTATGCACCTCTTGCTCACAATGGTCTGAATATGGGACTACATCACAATCTTT
TGTCTCGGCTTATTTAAGACAAAATCATTGGAATAACATAACTTACAGTACTTGTTTTTT
TTCTCGTTCTATATTCATGATTATGGTTATTTTTTGTGTTGAGTTGTTCAATTTTTCTGTC
TATTGCGTTCTATGAACCTTGACACTCTTTTTGTAAATTATTATATGCTAAAGACAATTTGG
ACTAACAGCATTTTTAATAAAAAAAAAAAAA

>G2148 Amino Acid Sequence (conserved domain in AA coordinates:130-268)

MENEAFVDGELESLGFMFNFDQCSSNESSFCNAPNETDVFSSDDFFPFGTILQSNYAVAL
DGSNHQTNRNVDNRQDLLKPRKKQKLSSESNLVTEPKTAWRDGQSLSSYNSSDDEKALGL
VSNTSKSLKRKAKANRGIASDPQSLYARKRRERINDRLKTLQSLVPNGTKVDISTMLEDA
VHYVKFLQLQIKLLSSEDLWMYAPLAHNGLMGLHNNLSRLI*

>G2379 (52..798)

CGCCGTCACCTCTCTCCCGTGCCGCACATTAGCAACACTACTCCCGACGAATGGAGACG
ACGACGCCGCAGTCAAAATCAAGTGTGTCCACCGACCGCCGTTGGGAAGAGAAGACTGG
TGGAGTGAGGAAGCGACGCGGACGCTGGTAGAAGCCTGGGGCAATCGTTACGTCAAGCTG
AACCACGGAATCTCCGGCAGAATGACTGGAAAGACGTCGCCGACGCCGTTAACTCTAGA
CACGGTGATAACAGCCGTAAGAAGACCGACTTACAGTGTAGAACCAGGGTCGATACTTTG
AAGAAGAAGTACAAAACAGAGAAAGCTAAACTCTCGCCGTCGACTTGGCGTTTCTATAAC
CGCCTCGATGTTCTAATCGGTCCCGTTGTGAAGAAATCGGCTGGCGGAGTTGTCAAATCA
GCGCCTTTTAAGAATCATCTGAATCCAACCTGGATCGAATCTACTGGAAGCTCTCTTGAA
GATGATGATGAGGATGATGATGAGGTTGGTGATTGGGAATTCGTTGCTAGGAAGCATCCT
CGTGTGGAAGAGGTAGATCTGAGTGAAGGATCAACGTGTAGGGAAGTACTACGGCGATT
CTCAAGTTTGAGAAGTTTACGAGAGAATTGAAGGGAAGAAGCAACAGATGATGATTGAG
TTGGAGAAGCAGAGAATGGAAGTGACAAAGGAGGTAGAGTTAAAACGAATGAACATGTTG
ATGGAGATGCAGTTAGAGATTGAGAAATCAAAGCACCGGAAACGCGCAAGTGCTTCAGGT
AAGAAGAAGTCAACATTAGG

>G2379 Amino Acid Sequence (domain in AA coordinates:19-110, 173-232)

METTTTPQSKSSVSHRPLRGREDWSEETATLVEAWGNRYVKLNHGNLRQNDWKDVADAV
NSRHGDNRSRKKTDLQCKNRVDTLKKKYKTEKAKLSPSTWRFYNRLDVLIGPVVKKSAAGV
VKSAPFKNHLPNTGNSNLTGSSLEDDDDDEVGDWFEFVARKHPRVEEVDLSEGSTCRELA
TAILKFGEVYERIEGKKQMMIELEKQRMVETKEVELKRMNMLMEMQLEIEKSKHRKRAS
ASGKKNSH*

>G1462 (63..1031)

CGTCGACCATTCTTGCGATTGATCTTTCTCTAGATAAATTTTTTGTATCGATTTAGTTTCA
TTATGGAGGACGACGACGACGCTTATGATCTAATCAAACACGAACGTGTATACCTCAGAAG
ACGAAGTAATAATCTCACGTTATCTGAAGGTTATGGTCGTTAACGGAGATTCTTGCCAG
ATCACTTTCATCGAAGACGCAACCGTGTTCACCAAGAATCCAGATAAGGTGTTCAATTCTG
AGAGACCTAGATTCTGTGATCGTTAAACCACGAACAGAGGCTTGTGGTAAAACCGATGGAT
GTGATTCGGGTTGCTGGAGGATCATTGGTCGTGATAAACTGATAAAGTCGGAGGAGACTG
GGAAGATTCTAGGGTTCAAGAAGATACTCAAGTTTTCCTAAAGAGGAAACCTATAGACT
ACAAGAGAAGTTGGGTAATGGAAGAGTATAGGCTTACCAATAACTTGAAGTGAAGCAAG
ATCATGTGATTTGCAAAATTCGTTTATGTTTGAAGCTGAAATTAGTTTCTTGCTAAGCA
AGCATTTCTACACTACATCAGAATCGGTTCTTGAAAATGAGCTGTTGCCATCTTATGGAT
ATTATTTATCCAATACACAAGAGGAGGATGAATTTATCTGGACGCGATAATGACTTCGG
AAGGAAACGAGTGGCCTAGCTACGTTACCAACAACGTGTACTGTCTGCATCCATTGGAGC
TTGTGGATCTTCAAGATCGGATGTTTAATGATTACGGAACCTGCATCTTCGCTAACAAGA
CTTGTGGTGAAAAGTATAAATGCGATGGTGGTTACTGGAAGATCCTGCACGGTGATAAGC
TGATCAAGTCAAATTCGGAAGGTCATTGGTTTCAAGAAGGTATTTGAGTTCTATGAAA
CGGTGAGACAAATATATCTTTGTGATGGAGAAGAAGTGACGGTAACCTGGACTATACAAG
AGTATAGGCTTAGCAAAAACGTGAAGCAGAATAAAGTGTGTGCGTTATCAAGTTGACTT
ATGATAGATAGGATACTTTACTTTGGTTTTTGTGATCATCTTAGTATCTTACGAATATTC
TAGATACACATCTATAGGCGACCGCTCTAGACAGGCCTCGTACCG

>G1462 Amino Acid Sequence (domain in AA coordinates: TBD)

MEDDDAAYDLIKHELLYSEDEVIIISRYLKGMVVNGDSWPDHFIEDANVFTKNPDKVFNSE
RPRFVIVKPRTEACGKTDGCDSGCWRIIGRDKLIKSEETGKILGFKKILKFCLKRKPIDY
KRSWVMEYRLTNNLNWKQDHVICKIRFMFEAEISFLLSKHFYTTSESVLENELLPSYGY
YLSNTQEBDEFYLDAIMTSEGNEWPSYVTNNVYCLHPLQLVDLQDRMFNDYGTICIFANKT
CGETDKCDGGYWKILHGDKLIKSNFGKVIKVFYFYETVRQIYLCDEEVTVTWTIQE

YRLSKNVKQNKVLCVIKLTIDR*

>G1211 (44..1120)

TGAAACCTAGATTCTGCAACTGAATTCCTAATTCGAAAAAGAAATGGAGGGTTCGTCGTC
GACGATAGCAAGGAAGACATGGGAAC TAGAGAACAGCATTCTAACAGTAGACTCACCTGA
TTCAACCTCCGACAACATCTTCTACTACGACGATACTTCACAGACTAGGTTCAGCAAGA
GAAACCGTGGGAGAATGATCCTCACTACTTTAAACGAGTCAAGATCTCAGCGCTCGCTCT
TCTTAAGATGGTGGTTCACGCTCGCTCTGGTGGTACAATTGAAATAATGGGTCTTATGCA
AGGTAAGACCGATGGTGATACTATCATTTGTTATGGATGCTTTTGCTTTACCAGTGGAAGG
TACTGAGACAAGGGTTAATGCTCAGGATGATGCTTATGAGTACATGGTTGAGTATTCACA
GACCAACAAGCTCGCGGGGCGGCTGGAGAATGTTGTTGGATGGTATCACTCTCACCTGG
ATATGGATGCTGGCTCTCCGGTATTGATGTTTCTACGCAGACGCTTAACCAACAGCATCA
GGAGCCATTTT TAGCTGTTGTTATTGATCCCACAAGGACTGTTTCAGCTGGTAAGGTGGA
GATTGGTGGCTTT CAGAACATACCTCTAAAGGATATAAGCCTCCAGATGAACCTGTTTCTGA
GTATCAAAC TATTCCTTTAAATAAGATTGAGGACTTTGGTGTTCACTGCAAACAGTACTA
TTCATTAGATGTCACCTTATTTCAAGTCATCTCTTGATTCTCACCTTCTGGATCTACTATG
GAACAAGTACTGGGTGAACACTCTTTCTTCTTCTCCACTGCTGGGTAATGGAGACTATGT
TGCTGGACAAATATCAGACTTAGCTGAGAAGCTTGAGCAAGCCGAGAGTCATCTGGTTCA
GTCTCGCTTTGGAGGAGTTGTGCCATCATCCCTTCATAAGAAAAAGAAAGATGAGTCTCA
ACTAACTAAGATAACTCGGGATAGCGCAAAGATAACTGTGGAACAGGTCCATGGACTAAT
GTCGCAGGTCAATAAAGATGAATTATCAACTCAATGCGTCAGTCCAACAACAATCTCC
CACTGACTCGTCCGATCCAGACCCCTATGATTACATATTGAAGTTGCTCTTCTTTTGGTTT
CTANTTTTGGATTGACCCATCATTTGTTGCTCTTTCATTTATTTTCTGTTGTGTAAGAA
TTATAATGNCNGCNGCAATTGCGGGCCGCTAAAAAANACAGGAAATTGAAAAAATTCN
NCCATTCCAACATCTTTATTTAATATTATCTCCTCNATTATATAATATTCAAACATCCCT
ANTANCTTCATTTGACCGTCCCCCTCCCTCCCGTGTGCTTGGTGTGGCCCC

>G1211 Amino Acid Sequence (domain in AA coordinates: 123-179)

MEGSSSTIARKTWELENSILTVDSPDSTSDNIFYDDTSQTRFQOEKPWENDPHYFKRVK
ISALALLKMWVHARSGGTIEIMGLMQKTDGDTIIVMDAFALPVEGTETRVNAQDDAYEY
MVEYSQTNKLAGRLENVVGWYHSHPGYGCWLSGIDVSTQTLNQHQEPFLAVVIDPRTV
SAGKVEIGAFRTYSKGYKPPDEPVSEYQTIPLNKIEDFGVHCKQYYSLDVTFYKSSLD
LLDLLWNKYWVNTLSSPLLNGNDYVAGQISDLAEKLEQAESHLVQSRFGGVVPSLHKK
KEDESQLTKITRDSAKITVEQVHGLMSQVIKDELFNMSRQSNKSPDSSDPDPMTY*

>G1048 (5..892)

GACCATGGCGGAGGAATTTGGAAGCATAGATTTACTCGGAGATGAAGATTTCTTCTTCGA
TTTCGATCCTTCAATCGTAATTGATTCCTTCCGGCGGAGGATTTCTTCAGTCTTCACC
GGATTCATGGATCGGAGAAATCGAGAATCAATTGATGAACGATGAGAATCATCAAGAGGA
GAGTTTGTGGAAATTGGATCAGCAATCGGTTTCAGATTTTCATAGCGGATCTACTCGTTGA
TTATCCAAC TAGCGATTCTGGCTCCGTTGATTGGCGGCTGATAAGTTCTAACCGTTCGA
TTCTCCCGCGCCGCTGATGATTCCGGGAAGGAGAATTCGGATTTGGTTGTTGAGAAGAA
GTCTAATGATTCCTGGTAGCGAGATTGATGATGATGACGAAGAAGGAGACGATGATGC
TGTGGCTAAAAACGAAGAAGGAGAGTAAGAAATAGAGATGCGCGGTTAGATCGAGAGA
GAGGAAGAAGGAATATGTACAAGATTTAGAGAAGAAGAGTAAGTATCTCGAAAGAGAATG
CTTGAGACTAGGACGTATGCTTGAGTGCTTCGTTGCTGAAAACCAAGTCTCTACGTTACTG
TTTGCAAAAGGGTAATGGCAATAATACTACCATGATGTCGAAGCAGGAGTCTGCTGTGCT
CTTGTTGGAATCCCTGCTGTTGGGTTCCCTGCTTTGGCTTCTGGGAGTAACTTCATTTG
CCTATTCCCTTATATGTCCACACAAAGTGTGCTCCTACGTCCAGAACCAGAAAAGCT
GGTTCTAAACGGGCTCGGGAGTAGTAGCAAACCGTCTTATACCGGCGTTAGTCGGAGATG
TAAGGGTTCGAGGCCTAGGATGAAATACCAAATCTTAACCTTTCGGGCGTGACAACGCCT
TTTTTAAGTCTTCTTTGCGCATTTTGAGTTGTAGATGAGTGTCTTTTAGTTTTCTCTC
TCTGTGTTTGTATTTCTGCTGTTGAAAGTTTCTGTCTAATATCGATAAGTTAACAGTGAA
AAAAAAAAAAAAAA

>G1048 Amino Acid Sequence (domain in AA coordinates 138-190)

MAEEFGSIDLLGDEDFPFDFDPSIVIDSLPAEDFLQSSPDWIGEIENQLMNDENHQEES
FVELDQQSVSDFIADLLVDYPTSDSGSVDLAADKVLTVDSPAAADDSGKENSIDLVEKKS
NDSGSEIHDDDDDEEGDDDAVAKRRRRVRNRDAVRSRERKKEYVQDLEKKSKYLERECL
RLGRMLECFVAENQSLRYCLQKGNNGNTTMMSKQESAVLLLESLLGSLWLLGVNFICL
FPYMSHTKCCLLRPEPEKLVNLGLSSSKPSYTGVSRRCKGSRPRMKYQILTLAA*

>G986 (31..846)

CATTAAATTGGCTCCTGTGAACCTAAATTTATGGACTATGATCCCAACACCAATCCGTTCT
GACCTTCATTTCTCCGGTAAACTTCCGAAAAGAGAAGTCTCGGCTTCAGCTTCTAAAGTT
GTAGAGAAGAAATGGTTAGTGAAAGATGAGAAGAGAAATATGCTACAAGATGAAATAAAC
CGGGTTAATTCCGGAGAACAAGAGCTAACCGAAATGTTAGCAAGAGTCTGTGAGAAGTAC
TATGCTCTTAATAATCTTATGGAGGAGTTGCAGAGTCGAAAGAGTCCTGAAAGTGTTAAC
TTTCAGAACAAACAGCTAACGGGGAAACGAAAACAAGAACTTGATGAGTTTGTTAGCTCC
CCAATTGGACTCAGTCTCGGACCAATCGAGAACATCACCAACGATAAAGCGACGGTTTCA
ACCGCTTACTTTGCTGCTGAGAAGTCTGACACAAGCTTGACTGTGAAAGATGGATATCAA
TGGAGGAAATACGGGCAAAAGATTACGAGAGATAATCCATCTCCTAGAGCTTACTTCAGA
TGCTCGTTTTTACCCTCTTGTCTAGTCAAGAAGAAGGTGCAACGAAGTGCAGAAGATCCA
TCTTTCTTGGTAGCCACTTACGAAGGGACACATAACCACACCGGACCACATGCAAGTGTG
TCCAGGACAGTGAAACTTGATCTAGTTCAAGGTGGGCTTGAACCAGTTGAGGAAAAGAAA
GAGAGAGGGACGATTCAAGAGGTTTGGTGCAACAAATGGCTTCTTCGTTGACCAAAGAT
CCTAAGTTCACTGCAGCTCTTGGCAGTCTATTTCCGGGAGATTGATAGAGCATTCAAGA
ACATGAAAGTTCTCTAGAACATGTATATTTCTGTTTTGTTCTATTTTGTGCTCATTCCT
AGTAAAAAGGTAAAGATTTGTTTGATCTTGATTAGGAGGCATAGATGTCAATTTTAATGT
GTGTGTATATAATTACATCAAACTAAGTATCCAAAAGGGTCACCCCATTTTATCTTA
TG

>G986 Amino Acid Sequence {domain in AA coordinates: 146-203}
MDYDPNTPNPFDLHFSGLPKREVSAASASKVVEKKWLKDEKRNMLQDEINRVNSENKKLT
EMLARVCEKYYALNNLMEELQSRKSPESVNFQNKQLTGKRKQELDEFVSSPIGLSLGPIE
NITNDKATVSTAYFAAEKSDTSLTVKDGQWRKYQKQITRDNPSPRAYFRCSFSPSCLVK
KKVQRSAEDPSFLVATYEGTHNHTGPHASVSRTVKLDLVQGGLEPVEEKKERGTIQEVLV
QQMASSLTKDPKFTAALATAISGRLIEHSRT*

>G789 (259..1593)

GGCAAGAAGAACCTTAGCCTCTCTTTCTTCTTTCTCTCTCTCTCTCTGTGGTACTGTT
CTGTTTCAACTTTACTCCCTCAGTTTCAGAACAAATCCCTATCTAGAAGAGAGATAAAAC
CGAGAAGGTTTGGAGATAGAATCTTTTGTCTTCTTTTGTCCCTCCTTGCTCGATTTTT
GTTACGTGTGAAGCAATAAAAAAACTGATATAGCTAAATCTTCCATCCATTGAGAGGC
TTCTAAATCTGATCTGACATGGAACAAGTGTTTGTGATTGGAATTTTGAAGATAATTTT
CACATGTCCACTAATAAAAGATCAATCAGACCAGAAGATGAATTAGTGGAGCTATTGTGG
AGAGATGGTCAAGTGGTTTTTACAAAGCCAAGCTCGTAGAGAACCGTCAGTCCAAGTCCAA
ACCCACAAACAAGAAACCCCTAAGAAAACCCAACAATATTTTCTTGACAACCAAGAAACA
GTACAAAAGCCTAACTACGCTGCTCTAGATGATCAAGAAAACCGTCTCCTGGATACAATAC
CCTCCGGATGACGTCATCGACCCCTTTCGAATCCGAGTTCTCCTCTCATTTCTTCTCTCG
ATCGATCACCTCGGAGTCTCTGAGAAGCCACGAACGATCGAAGAGACAGTTAAGCATGAG
GCTCAAGCCATGGCTCCTCTAAGTTTAGATCCTCGGTTATAACAGTCGAGCCGAGTCAT
TGCGGCAGCAACCAGTCAACAAATATTATCAGGCCACTACACTTCCGGTTTCTATGAGT
GATAGAAGCAAGAACGTCGAAGAAAGACTTGACACTTCGTCAGGTGGCTCCTCCGGTTGC
AGCTATGGAAGGAACAACAAGAAACCGTTAGTGGAACAAGTGTAACCATGACCGTAAA
AGAAAACATGTTATGGATGCTGATCAAGAATCTGTGTCTCAATCAGATATAGGTTTGACC
TCAACCGATGATCAAACCATGGGTAACAAATCGAGCCAACGGTCAGGATCTACTCGAAGA
AGCCGTGCAGCTGAAGTTTCAATCTCTCAGAAAGGAGGAGAGATCGGATCAATGAA
AGAATGAAAGCTCTTCAAGAACTCATACCTCACTGCAGCAGAACAGATAAAGCTTCGATA
TTGGATGAAGCAATTGATTACTTAAATCACTTCAAATGCAACTCCAAGTGATGTGGATG
GGAAGTGAATGGCGGCGGCAGCAGCAGCAAGTCCGATGATGTTCCCGGGGTA
CAATCATCTCCATACATTAATCAGATGGCTATGCAAAGTCAGATGCAATTGTCTCAATTC
CCGTTATGAACCGGTCCGCTCCGAGAACCATCCCGGTTTAGTATGTCAAAACCCGGTA
CAGTTGCAGCTCCAAGCACAGAACCAATCTTATCGGAGCAGCTCGTAGGTACATGGGC
GGGATTCCCCAGATGCCGCCGCGGGAAATCAGATGCAGACCGTGCAACAACAACAGCG
GACATGTTGGGATTTGGATCTCCGGCGGGACCGCAAAGTCAACTGTCCGCACCGGCGACC
ACCGACAGTCTTCATATGGGTAAAATAGGCTGACTTGGCATATAGTTTTCTCCGAAATT
ATCTTCTTACAGTTGGTGATGTTATTTATTTTGGTGCCTTAAGCAAGCATAAAAGCT
AAGTCAAATGTATTATAGAGATCTAATAAGTTAGTCTCATACTTATAACTTATTTTAA
CAGTTGAATTATAGTATCAATCAAGTGTTGGGAACCTAAAGATCATACATGTGTCAATAC
TTTTATATTTGTTCTCAAGGTTTCATCAGAAAAACAAAATAAAAAGGATAGACTAGGCCTG

CATTTGACATTATCATGGGCTTTTTTGGGTCTATGAATATGAACATTAACCCC
>G789 Amino Acid Sequence (domain in AA coordinates: 253-313)
MEQVFADWNFEDNFHMSTNKR SIRPEDELVELLWRDQGVVLQSQARREPSVQVQTHKQET
LRKPNNIFLDNQETVQKPNYAALDDQETVSWIQYPPDDVIDPFSEFSSHFFSSIDHLGG
PEKPRTIETVKHEAQAMAPPKFRSSVITVGPSHCGSNQSTNIHQATTLPVMSDRSKNV
EERLDTSSGGSSGCSYGRNNKETVSGTSTIDRKRKHVMDADQESVSQSDIGLSTDDQT
MGNKSSQSRSGSTRRSRAAEVHNLSERRRRDRINERMKALQELIPHCSRDKASILDEAID
YLKSLQMLQVMMWMSGMAAAAAAASPMMPGVQSSPYINQMAMQSQMQLSQFPVMNRS
APQNHPLVCQNPVQLQLQAQNLSEQLARYMGGIPQMPPAGNQMQTVQQQPADMLGFG
SPAGPQSQLSAPATTDLSLHMGKIG*
>G2085 (1..930)
ATGTTTGGTCGCCATTTCGATTATCCCAAATAACCAGATTGGTACCGCCTCTGCTTCCGCT
GGTGAAGACCATGTCTCTGCCCTCCGCTACGTCTGGTCACATTCCCTTACGACGATATGGAA
GAAATCCCTCATCTGACTCTATCTATGGTGCTGCCCTCCGATTTGATTCCCGATGGCTCT
CAATTGGTTGCTCACCGATCCGATGGCTCTGAATTACTTGTCTCTCGGCCACCGGAAGGG
GCGAATCAGCTTACGATCTCGTTCCTGGACAAGTTTACGTTTTTGTATGCCGTGGTGCT
GACAAGGTGGATGCTGTGTTGTCTGCTGTTGGGTGGTCTACTGAGCTTGCTCCTGGTCCG
CAGGTGATGGAAGTAGCTCAACAGCAGAATCATATGCCTGTTGTAGAAATATCAGAGCCGC
TGTAGCCTTCCGCAACGGGCACAATCCTTGGATAGGTTTCGGAAGAAGAGGAATGCTAGA
TGTTTCGAGAAGAAAGTAAGATACGGTGTTCCGCAAGAAGTTGCCTTAAGAATGGCACGT
AATAAAGGTCAATTACCTCTTCAAAGATGACAGATGGGGCTTATAACTCTGGCACAGAT
CAAGATTCGCCAAGATGATGCCCATCCAGAAATATCGTGTACTATTGCGGCATTAGT
TCCAAATGTACACCAATGATGCGACGTGGCCCTTCCGGCCCCAGGACTCTCTGCAATGCC
TGTGGACTTTTTTGGGCTAACAGGGGTACATTGAGGGATCTCTCAAAGAAAACAGAAGAG
AATCAGTTGGCTTTAATGAAACCGGATGATGGTGGGAGTGTGCTGATGCTGCTAACAAC
TTAAACACTGAAGCTGCAAGTGTGTAAGAACACACTTCCATGGTTTTCTTGTCCAATGGG
GATAATTCTAATCTGTTAGGTGATCACTAA
>G2085 Amino Acid Sequence (domain in AA coordinates: TBD)
MFGRHSIIPNNQIGTASASAGEDHVSASATSGHIPYDDMEIIPHPDSIYGAASDLIPDGS
QLVAHRSDGSELLVSRPPEGANQLTISFRGQVYVFDVAVGADKVDVAVLSLLGGSTELAPGP
QVMELAQQNHPVVEYQSRCSLPQRAQSLDRFRKRNARCFEKKVRYGVRQEVALLRMAR
NKGQFTSSKMTDGAYNSTGTDQSDAQDDAHPEISCTHCGISSKCTPMMRRGPGSPRTLNA
CGLFWANRGTLRLDLKKTEENQLALMKPDDGGSVADAANNLNTEAASVEEHTSMVSLANG
DNSNLLGDH*
>G1783 (1..603)
ATGGCCGCGTTTCCGCAAGGGTCGATGACAAACGTTTTGAGTTAGCTCTGCTT
CAAATCCCGAGGGTTCGCCGAATTTATAGAGAATATCGCCTATTATCTCCAGAAACCG
GTGAAGGAGGTGGAGTACTACTGCGCGTTGGTCCATGATATTGAGCGGATCGAATCG
GGTAAGTATGTTTTGCCCAAATACCCGGAAGACGATTACGTGAAACTGACGGAAGCAGGT
GAGTCTAAGGGCAATGGGAAAAAGACGGGAATTCCTTGGTCAGAAGAGGAACAGAGGTTG
TTTCTGGAAGGACTAAATAAGTTTTGGGAAAGGAGACTGGAAGAACATATCGAGGTATGT
GTGAAGTCAAGGACCTCGACGCAAGTGGCAAGCCATGCTCAGAAGTATTTTGCAAGGCAA
AAGCAGGAGAGTACGAATACTAAACGCCCCGAGTATTCATGACATGACTCTGGGAGTTGCG
GTCAATGTCCCTGGATCCAACCTGGAGTCTACTGGCCAGCAACCACATTTTGGTGATCAA
ATTCTTTCGAATCAATATTATCCCTCCAGGAAAACCTTTCGGGGTTTTGATCAGCGATGG
TGA
>G1783 Amino Acid Sequence (domain in AA coordinates: 81..129)
MAAFPQWTRVDDKRFELALLQIPEGSPNFIEINAYYLQKPVKEVEYYYCALVHDIERIES
GKYVLPKYPEDDYVKLTEAGESKNGKKTGIPWSEEEQRLFLEGLNKF GKGDWKNISRYC
VKSRTSTQVASHAQKYFARQKQESTNTKRPSIHDMTLGVAVNVPGSNLESTGQQPHFGDQ
IPSNQYYPSENFRGFDQRW*
>G2072 (155..793)
TCGACCCACGCGTCCGCCACGCGTCCGGATCTTTTCACAGAAGACCAACCAGCTTGGCT
CGATGAGCTCCTAAGTGAGCCAGCATCACCTAAGATTAACAAAGGTCATAGACGTTCCAGC
TAGTGACACAGCTGCTTACTTGAAGTCAAGCTTTAATGCCTTCGAAGGAAAATCATGTTGC
TGGTTCGTCTTGGCAGTTCCAGAACTATGATTTGTGGCAGTCCAACCTTATGAACAACA
CAATAAATTAGGATGGGATTTCTCTACAGCAAATGGAATAATATCAAAGAAATATGTC

ATGCGGAGCTTTAAATATGTCGTCGAAACCCATTGAGAAACATGTAAGCAAAATGAAAGA
AGGAACCTTCTACAAAACAGATGGTCTTAGATCAAAGACTGACTCAAACGTATCAAACA
TCAAAATGCTCATCGAGCGCGTTTGAAGGCTTGAGTACATATCAGACCTTGAAAGGAC
CATCCAAGTGCTACAAGTTGAAGGATGTGAAATGTCATCTGCCATTCACTACTTGGATCA
GCAGTTACTCATGCTTAGCATGGAAAAATAGAGCTTTAAAAACAACGTATGGATAGTTTAGC
AGAAATCCAAAAGCTTAAACATGTGGAGCAGCAATTGCTTGAGAGAGAGATAGGAAACCT
ACAGTTTCGACGACACCAACAACAACCACAGCAAAACCAAAAACAAGTCCAAGCAATACA
AAATCGATACACCAAAATATCAACCACCTGTTACACAAGAACCCGATGCCAATTTGCAGC
CTTGGCAATATGATTTAGGAAATATGGATACATTGTTTCAGATTAAGCTGAGCTCCTCTTG
CTCTACCTTAATGTCCATACAAACATAGGTGAACCTTGATGTTTGTAGCCTTGAATGAAAC
CTAAAAAGCATCGTTATGTAAATCAAATGTGGTTGCCCATATCCTCCTCTATTGCATT
TCTCTCTATTATGGCATGGTAGAGAATCTCTTGTCAAGAACTTCATGTTATGTAATAA
CTTGTAACTCTTCTTATTCATCTATTATATATATGAATAAGTAATTTTTTTGCCAAAAA
AAAAAAAAAAAAAAAAAAAA

>G2072 Amino Acid Sequence (conserved domain in AA coordinates:90-149)

MPSKENHVAGSSWQFQNYDLWQSNSEYQHNKLGWDFSTANGTNIQRNMSCGALNMSSKPI
EKHVS KMKEGTSTKPDGPRSKTDSKRIKHQNAHRARLRRLLEYISDLERTIQVLQVEGCEM
SSAIHYLDQQLMLSMENRALKQRMDSLAEIQKLKHVEQQLLEREIGNLQFRRHQQPQQ
NQKQVQAIQNRYYTKYQPPVTQEPDAQFAALAI*

>G931 (85..1071)

GGAGGTTCTTTGACAGACACATGTATCATCAATCTTCTCTGTTGAAGCAGAGAGAGAGAG
AGCTAATTGTTGCCTCTGAGTCACATGGATAAGAAAGTTTCATTTACTAGCTCTGTGGCA
CATTCAACTCCACCATACTTAGTACTTCCATCTCATGGGGACTTCCAACCAAATCCAAT
GGTGTGACTGAATCACTGAGTTTGAAGGTGGTAGATGCAAGACCAGAACGTCTTATAAAC
ACAAAGAAATACAGTTTCCAGGACCAGGATTCATCTTCAACTCTGTCTCTGCTCAATCT
TCTAACGATGTTACAAGTAGTGGAGATGATAACCCCTCAAGACAAATCTCATTTTTAGCA
CATTCAAGATGTTGTAAAGGATTTGAAGAACTCAAAGGAAGCGATTGCAATTAAATCA
GGCTCCTCCACGGCAGGAATCGCTGATATTCACTCTTCTCCTTCCAAGGCTAACTTCTCA
TTTCTGATATGCCGATCCACATTTTGGTGGTTTAAATGCCTGCGGCTTACCTACCACAGGCA
ACAATATGGAAATCCCAAATGACTCGAGTTCCGCTACCATTCGATCTCATAGAGAATGAG
CCTGTCTTTGTCAATGCAAAGCAATTCCATGCAATTATGAGGAGGAGGCAACAGCGTGCT
AAGCTAGAGGCGCAAAACAACTAATCAAAGCCCGTAAGCCGTATCTTCATGAATCTCGA
CATGTTACAGCTCTTAAACGACCTAGAGGATCTGGTGAAGATTCTTAAACACCAAAAAG
CTTCAAGAAATCTACAGATCCAAAACAAGACATGCCAATCCAACAGCAACACGCAACGGGA
AACATGTCAAGATTTGTGCTTTATCAGTTGCAGAACAGCAATGACTGTGATTGTTCAACC
ACTTCTCGCTCTGACATCACATCTGCTTCTGACAGCGTTAATCTCTTTGGACACTCTGAA
TTTCTGATATCAGATTGCCCATCTCAGACAAACCAACAATGTATGTTTCATGGTCAATCA
AATGACATGCATGGAGGTAGGAACACACACCATTCTCTGTCCATATCTGAGCCGGTGGA
ATCTGGTAATGTGTACGTTCTTACAAAAAAGGGAAGTCATCCTTGGCTGCTACTTCGCT
TATTAGCTAGTCTTATTTTACACGCTTTGTCCAGATATC

>G931 Amino Acid Sequence (domain in AA coordinates: TBD)

MDKKVSFTSSVAHSTPPYLSISWGLPTKSNVTESSLKLVVDARPERLINTKNISFQD
QDSSSTLSSAQSSNDVTSSGDDNPSRQISFLAHSVDCKGFEETQKRFAIKSGSSTAGIA
DIHSSPSKANFSFHYADPHFGLMPAAYLPQATIWNPMQTRVPLPFDLIENEPVFNKQ
FHAIMRRRQQRKLEAQNKLKARKPYLHESRHHVHALKPRGSGGRFLNKKLQESTDPK
QDMPIQQQHATGNMSRFVLYQLQNSNDCDCSTTSRSDITSASDSVNLFGHSEFLISDCPS
QTNPTMYVHGQSNMDMHGGRNTHHFSVHI*

>G278 (93..1874)

TCGATCTTTAACCAATCCAGTTGATAAGGTCTCTTCGTTGATTAGCAGAGATCTCTTTA
ATTTGTGAATTTCAATTCATCGGAACCTGTTGATGGACACCACCATGATGGATTGCGCG
ATTCTTATGAAATCAGCAGCACTAGTTTCGTCGCTACCGATAACACCGACTCCTCTATTG
TTATCTGCGCCCGCAACAAGTACTCACCGGACCTGATGTATCTGCTCTGCAATTGCTCT
CTTAACAGCTTCGAATCCGTCTTTGACTCGCCGGATGATTTCTACAGCGACGCTAAGCTTG
TTCTCTCCGACGCGCGGAAGTTTCTTTCCACCGGTGCGTTTTGTACGCGAGAAGCTCTT
TCTTCAAGAGCGCTTTAGCCGCCGCTAAGAAGGAGAAAGACTCCAACAACACCGCGCGCG
TGAAGCTCGAGCTTAAGGAGATTGCCAAGGATTACGAAGTCGGTTTTCGATTGCGTTGTGA
CTGTTTTGGCTTATGTTTACAGCAGCAGAGTGAGACCGCCGCTAAAGGAGTTTCTGAAT

CGCGAGACGAGAATTGCTGCCACGTGGCTTGCCGGCCGGCGGTGGATTTCATGTTGGAGG
TTCTCTATTTGGCTTTTCATCTTCAAGATCCCTGAATTAATTACTCTCTATCAGAGGCACT
TATTGGACGTTGTAGACAAAGTTGTTATAGAGGACACATTGGTTATACTCAAGCTTGCTA
ATATATGTGGTAAAGCTTGTATGAAGCTATTGGATAGATGTAAAGAGATTATTGTCAAGT
CTAATGTAGATATGGTTAGTCTTGAAGAGTCATTGCCGGAAGAGCTTGTAAAGAGATAA
TTGATAGACGTAAAGAGCTTGGTTTGGAGGTACCTAAAGTAAAGAAACATGTCTCGAATG
TACATAAGGCACTTGACTCGGATGATATTGAGTTAGTCAAGTTGCTTTTGAAAGAGGATC
ACACCAATCTAGATGATGCGTGTGCTCTTCATTTGCTGTTGCATATTGCAATGTGAAGA
CCGCAACAGATCTTTTAAACTTGATCTTGCCGATGTCAACCATAGGAATCCGAGGGGAT
ATACGGTGCTTCATGTTGCTGCGATGCGGAAGGAGCCACAATTGATACTATCTCTATTGG
AAAAAGGTGCAAGTGCATCAGAAAGCAACTTTGGAAAGGTAGAACCAGCACTCATGATCGCAA
AACCAAGCCACTATGGCGGTTGAATGTAATAATATCCCGAGCAATGCAAGCATTCTCTCA
AAGGCCGACTATGTGTAGAAATACTAGAGCAAGAAGACAAACGAGAACAAATTCCTAGAG
ATGTTCCCTCCCTCTTTTGCAGTGGCGGCCGATGAATTGAAGATGACGCTGCTCGATCTTG
AAAATAGAGTTGCACCTTGCTCAACGTCTTTTTCCAAACGGAAGCACAAGTGCAATGGAGA
TCGCCGAAATGAAGGGAACATGTGAGTTCATAGTGAAGTCTCGAGCCTGACCGTCTCA
CTGGTACGAAGAGAACATCACCGGGTGTAAGATAGCACCTTTCAGAATCCTAGAAGAGC
ATCAAAGTAGACTAAAAGCGCTTTCTAAAACCGTGGAACCTCGGGAACGATTCTTCCCGC
GCTGTTCCGCGAGTGTCTGAGAACGACTACAAAAGAAAGCAAGGTACATGAAATACAAG
AGACACTAAAGAAGGCCTTTAGTGAGGACAATTTGGAATTAGGAAATTCGTCCCTGACAG
ATTGCACTTCTTCCACATCGAAATCAACCGGTGGAAGAGGTCTAACCGTAAACTCTCTC
ATCGTCTGTCGGTGAGACTCTTGCTCTTAGTGTAATTTTGTGTACCATATAATTCTGT
TTTCATGATGACTGTAAGTGTATGTCTATCGTTGGCGTCATATAGTTTCGCTCTTCGT
TTTGCATCCTGTGTATTATTGCTGCAGGTGTGCTTCAAACAAATGTTGTAACAATTTGAA
CCAATGGTATACAGATTGTGAATATATATTTATGTACATCAACAATAAAAAAAAAAAAAA
AAAA

>G278 Amino Acid Sequence (domain in AA coordinates: 2-593)
MDTTIDGFADSYEISSTSFVATDNTDSSIVYLAEEQVLTGPDVLSALQLLSNSFESVFDSP
DDFYSDAKLVLSDBGREVSFHRVLSARSSFFKSALAAKKEKDSNNTAAVKLELKEIAKD
YEVGFDSVVTVLAYVYSSRVPPKGVSECADENCCHVACRPVDFMLEVLYLAFIFKIP
ELITLYQRHLLDVVDKVVIEDTLVILKLANICGKACMKLLDRCKEIIIVKSNVDMVSLEKS
LPEELVKEIIDRRKELGLEVPKVKHVS NVHKALDSDDIELVKLLKEDHTNLDDACALH
FAVAYCNVKTATDLLKLDLADVNHRNPRGYTVLHVAAAMRKEPQLILSLLEKGASASEATL
EGRTALMIKQATMAVECNNIPEQCKHSLKGRLCVEILEQEDKREQIPRDVPPSFAVAAD
ELKMTLLDLBNRVALAQRLLFPTAQAAAMEIAEMKGTCEFI VTSLEPDRLTGTRKTS PGVK
IAPFRILEEHQSRLKALS KTVELGKRFFPRCSAVLDQIMNCEDLTQLACGEDDTAEKRLQ
KKQRYMEIQETLKKAFSEDNLELGNSSLTDSTSSTSKSTGGKRSNRKLSHRRR*

>G2421 (1..630)

ATGGAGGGTTCGTCCAAAGGGTTGAGGAAAGGTGCATGGACTGCTGAAGAAGATAGTCTC
TTGAGGCAGTGATTTGGTAAGTATGGAGAAGGCAATGGCATCAAGTTCCTTTAAGAGCT
GGGCTAAATCGGTGCAGGAAAAGTTGTAGACTAAGATGGTTAAACTATTTGAGCCAAGT
ATCAAGAGAGGAAAATTTAGTTCTGATGAAGTTGATCTTCTTCTCGTCTTCATAAGCTT
CTAGGAAATAGGTGGTCTTGATTGCTGCTGATTACCTGGTCCGACCGCTAATGATGTC
AAGAACTACTGGAACACCCATCTGAGTAAGAAGCATGAACCGTGTGTAAACTAAGATA
AAAAGGATAAAATATTATAACCCCTCCTAATACACCGGCCCAAAAAGTTTGTGAAAATAGT
ATCACATGTAACAAAGATGATGAGAAAGATGATTTTGTGGATAATTTTATGGTTGGAGAT
AATATATGGTTGGAGCGTTTGTCTAGACGAGGGCCAAGAGGTAGATGTGCTGGTTACAGAA
GCGGCGGCAACAGAAAAGGAGGGCACTTTGGCGTTTGCAGTTGAGCAACTTTGGAATTTG
TTCGATGGAGAGACTGTGATCTTTGATTAGTGTATATAACGTTTGTGTTCTCTTGTGTTG
TGAGGTTTCTCTATTTAATTTAGTATCTATTTTCTAAATTAACATAATATCTTATAGTATT
TTAGGCAAACTTATGTTTCCGTTTCTGTCGCGCCGCTCTAG

>G2421 Amino Acid Sequence (domain in AA coordinates: 9-110)
MEGSSKGLRKGAWTABEDSLLRQCIGKYEGEKWHQVPLRAGLNRCRKSCRLRWLNLYLKPS
IKRGKFSSDEVDLLLRHLKLLGNRWSLIAGRLPGRANDVKNYWNTHLSKKHEPCCKTKI
KRINIIPTNPAPQKVCENSITCNKDEKDDFVDNFMVGDNIWLERLLDEGQEVLDVLVTE
AAATEKEGTLAFDVEQLWNLF DGETVIFD*

>G2032 Amino Acid Sequence (domain in AA coordinates: entire protein)

>G1396 (83..313)

>G1396 Amino Acid Sequence (domain in AA coordinates: TBD)

>G619 (382..2748)

ATTTTTTCCAATCTGCAAATTTTAGTCTATGTCTGTTCCCTTGTGCTCCCTCTTCTCAGT

ACCTGCAAATGGAGGAAGAAGAAATCCTTCTCTGAAACCCCTGTTCTCATTGATTCCTCC
TTCTCTCTCTTCTTCTCTCTCTGTCTCTGATTTCGTTATTCCACACTTATGACTCATCTT
TCCCGTCAATAGCTAAGTTTGCCCTCTTCTTTGTGAAATTTAGCTGAAAAAGGAGAGGAAT
TCCGAATTCGTCACTTCAAAGCTCGAATTTTGCAACTTTCTTTGATGGGTTTTACTT
GTTTTGTGTGAATCTGATTAAAAATAGAACTTTTTGTCTTTCTTCTGTCTCCTTTTGCT
CTTAAAGAGAAGCTTTTTCAATGGAATTTGACTTGAATACTGAGATTGCGGAGGTGGAA
GAGGAGGAGAATGATGATGTAGGAGTAGGAGTAGGAGGAGGAACAAGAATTGACAAGGGT
AGGCTTGGAATTTACCATCTTCTTCTTCTTCATGCTCTTCCGGATCATCATCGTCATCA
TCTTCTACAGGCTCTGCATCTTCCATTTACTCTGAGCTTTGGCATGCTTGTGCTGGTCCT
CTCACTTTCTTCCCAAGAAAGGCAATGTAGTTGTCTATTTCCCTCAAGGTCATTTGGAG
CAAGATGCTATGGTTTCATATTTCGTCTCCTCTTGAAATCCCCAAATTTGACCTTAATCCC
CAAATCGTCTGCAGGTTGGTTAATGTCCAGTTGCTTGCTAATAAGGACACCGATGAGGTC
TACACTCAAGTCACTGCTTCCACTTCAAGAGTTTTTCGATGCTAAATGGGGAGGGGAAA
GAGGTCAAGGAGTTAGGAGGGGAGGAAGAGAGGAACGGAAGCTCATCCGTCAAGCGGACA
CCTCATATGTTCTGTAAAACCTTAACAGCGTCTGACACAAGCACACATGGAGGCTTCTCT
GTACCTAGAAGAGCCGCTGAAGATTGTTTTGCTCCTCTTGACTACAAACAACAGAGGCCA
TCTCAAGAGCTCATTGCAAAGGACCTCCATGGAGTAGAGTGGAAGTTTCGCCATATCTAT
AGAGGTCAACCAAGGAGGCATCTACTCACCACCTGGTTGGAGTATCTTTGTCAGTCAAAAG
AATCTCGTCTCTGGTGATGCGGTTCTCTTTCTGAGAGACGAAGGAGGAGAGCTGAGATTA
GGAATCAGAAGAGCAGCAGCGCAAGAAATGGACTTCTGACTCAATCATGAGAAGAAT
TCATGTTCAAACATTCTGTCTCTTGTGGCTAATGCTGTATCTACAAAAAGCATGTTTCAT
GTGTTCTACAGTCCACGAGCGACGCATGCAGAGTTTGTGATTCTTATGAGAAGTATATC
ACAAGCATCAGGAGTCTGTTTGCATAGGCACAAGATTTAGAATGCGATTTGAAATGGAC
GATTCTCCTGAGAGAAGATGCGCTGGTGTAGTGACTGGAGTCTGTGACTTGGACCCGTAT
AGGTGGCCAACTCTAAATGGAGGTGCTTGTGTTGCGATGGGATGAGTCTTTGTGAGT
GATCATCAAGAAAGAGTTTCACTTGGGAGATTGATCCCTCGGTTTCTCTCCCACTTG
AGCATTCACTCATCTCAAAGGCTTAAAGGCCATGGGCAGGTTTACTGGATACTACCCCA
CCCGAAACCCCATAAACAAAAGGGGTGGTTTTTTGGACTTTGAGGAGTCGGTTAGACCC
TCTAAGGTCTTGCAAGGTCAAGAAAATATAGGTTCTGCATCACCCCTCACAGGGGTTTGAT
GTTATGAACCGCCGGATCTGATTTTGCATGCAGTCTCATGCAAATCCAGTCTTTGTG
TCGAGTAGAGTCAAGGATCGATTTGGTGAGTTTGTAGATGCTACTGGCGTGAACCCAGCT
TGTTTCAGGTGTTATGGACCTGGATAGGTTTCCAAGGGTCTTGCAAGGTCAAGAAATTTGC
TCGCTTAAATCATTCCCGCAATTTGCTGGTTTCAGTCCAGCTGCTGCTCCTAATCCCTTT
GCTTACCAAGCCAACAAGTCAAGTTACTATCCGCTAGCTTTGCATGGGATTAGGAGCACT
CATGTTCCGTATCAGAATCCATACAATGCGGGAAACCAATCCTCGGGTCCCCCTTCACGT
GCAATAAACTTTGGTGAAGAGACTAGAAAGTTTGATGCACAAAATGAAGGTGGCCTACCA
AATAATGTTACAGCTGATTTGCCATTCAAGATTGATATGATGGGAAAACAGAAAGGCAGT
GAGTTGAATATGAATGCTTTCATCAGGATGTAACTTTTTCGGATTCTCCTTACCAGTGGAG
ACACCTGCATCTAAGCCGCAAGCTCGAGCAAAAGAATCTGTACAAAGGTTCAAGCAA
GGAAGCCAAGTGGGAGAGCTATTGATTTGTGCGGACTTAACGGGTATGATGATCTCCTT
ATGGAGCTTGAACGGCTGTTCAACATGGAAGGGCTTCTCAGGGATCCTGAAAAGGATGG
AGGATCTTATATACTGATAGTGAGAACGATATGATGGTTCGTTGGCGATGATCCATGGCAT
GATTTCTGCAATGTGGTGTGGAAGATACACTTATACACGAAAGAGGAAGTGGAGAATGCG
AATGACGATAACAAGAGTTGTTTAGAGCAAGCTGCTCTCATGATGGAAGCATCAAAGTCA
TCTTCTGTGAGCCAGCTGATTTCTCTCCTACAATCACTAGGGTTTGATACCCATAAAGA
AGCTTATTTCTATGTTTTTAAAGTGTGTTTTGCTCACAAAAGAACTTCACTTTATCTTT
GTCTTTGAATCCATTTATGTGTTTGTGTTTCTTCTGGTCTCCATGGATGTCTCATG
TGTACCGTTTTACTGAGAGATATGTGAGTTTATGGGATGTGTAAAGCATGCCATTGGAT
TTTAAGGTTTTCAAATTACAAATATATATATTAGTTTTGAAGTTAAAAA

A

>G619 Amino Acid Sequence (domain in AA coordinates: 64-406)
MEFDLNTIEAEVEEENDDVGVGVGGGTRIDKGRLGISPSSSSSSSSGSSSSSSSTGSAS
SIYSELWHACAGPLTCLPKGNVVVYFPQGHLEQDAMVSYSSPLEIPKFDLNPQIVCRVV
NVQLLANKDTEVYTVQVTLPLQEFMSLNNEGKEVKELGGEERNGSSSVKRTPHMFCKT
LTASDTSHTGGFSVPRRAEDCFAPLDYKQQRPSQELIAKDLHGVEWKFRHIYRGQPRRH
LLTTGWSIFVSQKNLVSGDAVLFLRDEGGELRLGIRRAARPRNGLPDSIIKNSCSNLS
LVANAVSTKSMFHVFSRATHAEFVIPYEKYITSIRSPVICGTRFRMRFEMDDSPERRC

AGVVTGVCDLDPYRWPNKSWRCLLVWRWDESFVSDHQERVSPWEIDPSVSLPHLSIQSSPR
PKRPWAGLLDTPPGNPITKRGGFLLDFEESVRPSKVLQGGQENIGSASPSQGFVDMNRRIL
DFAMQSHANPVLVSSRVKDRFGEFVDATGVNPACSGVMDLDRFPRVLQGGQEICSLKSFQ
FAGFSPAAAPNPFAYQANKSSYYPLALHGIRSTHVPYQNPYNAGNQSSGPPSRAINFGEE
TRKFDAQNEGGLPNNVTDLPFKIDMMGKQKGSSELMNASSGCKLFGFSLPVETPASKPQ
SSSKRICTKVHKQGSQVGRAIDLRLNGYDDLLMELERLNFMEGLLRDPEKGWRILYTD
ENDMMVVGDDPWDFCNVVKIHLTYTKEEVENANDDNKSCLEQAALMMEASKSSSVSQPD
SSPTITRV*

>G2295 (33..917)

GTAATATATAACAATAACTCAGGTTACAAAGGATGGTTCGAAAGTGGTCGACCTACAAA
GGATAGCGAACGATAAGACAAGGATAACAACCTTACAAGAAGAGGAAAGCTAGTCTTTACA
AGAAGGCACAAGAGTTCTCAACTCTCTGCGGCGTCGAGACATGTCTCATCGTCTACGGTC
CCACGAAGGCTACCGATGTGGTGATTTCCGAGCCAGAGATATGGCCGAAGGACGAGACCA
AAGTCAGGGCCATCATACGCAAGTACAAAGACACAGTGTGACCAGCTGCAGGAAAGAAA
CCAACGTGGAGACTTTCGTCAACGATGTAGGGAAGGAAACGAGGTGGTGACTAAAAAGA
GAGTGAAGCGTGAGAATAAGTATTCTAGTTGGGAGGAGAAGCTAGACAAGTGTTCACGAG
AGCAACTACATGGGATTTTCTGTGCCGTGGATAGCAAGTTAAATGAAGCTGTAACGAGAC
AGGAGCGTAGTATGTTTAGGGTTAATCATCAAGCCATGGACACACCATTTCCCGCAGAATT
TAATGGACCAACAATTCATGCCACAGTATTTTCATGAGCAGCCACAGTTTCAAGGCTTCC
CTAATAATTTCAATAATATGGGTTTCTCGTTGATTTACCTCATGATGGTCAGATTCAA
TGGACCCAAATCTCATGGAGAAGTGGACCGACTTGGCTTTGACTCAAAGCTTGATGATGT
CAAAGGGAACGATGGTACTCAATTCATGCAGAGGCAAGAACAACCATACTATAATCGTG
AACAGGTTGTATCGAGGTCTGCAGGTTTCAATGTTAACCCGTTTATGGGATATCAAGTCC
CGTTTAATATTCTAATTGGAGATTATCGGGAATCAAGTTGAAAATTGGGAGCTTTTCAG
GGAAGAAAACGATATGATTTGAATTACGGAGCTTTATTAGTTTTTAGGGTTTTATAGTTT
TG

>G2295 Amino Acid Sequence (domain in AA coordinates: TBD)

MVPKVVDLQRIANDKTRITTYKKRKASLYKKAQEFSTLCGVETCLIVYGPTKATDVVISE
PEIWPKDETKVRAIRKYKDTVSTSCRKETNVETFVNDVGKNEVVTKKRVKRENYSSW
EEKLDKCSREQQLHGIFCAVDSKLNEAVTRQERSMFRVNHQAMDTPFPQNLMDQQFMPQYF
HEQPQFQGFNNFNMGFSLISPHDQIQMDPNLMEKWTDLALTQSLMMSKGNDDGTQFMQ
RQEQPYNREQVVSRSAGFNVNPFMGYQVFPNIPNWRLSGNQVENWELSGKKTII*

>G312 (1..1755)

ATGGCTTACATGTGCACTGATAGTGGCAATCTAATGGCTATTGCTCAACAAGTCATCAAA
CAGAAGCAGCAACAAGAACAACAACAGCAGCAACATCATCAAGACCATCAGATTTTGGT
ATTAATCCTTTGTCTCTTAACCCATGGCCCAATACTTCCCTCGGGTTTGGGCTTTCAGGT
TCGGCTTTTCCCGACCCGTTTCAAGTTACCGGCGGCGGAGATTCCAACGATCCTGGCTTT
CCTTTTCTCAACTAGACACCACCCACGCCACAACCCGCGGTGGGTTTCAGGTTATCT
GATTTTCGGCGGTGGAACCGGCGGCGGCGAGTTTGAGTCCGACGAGTGGATGGAGACTCTT
ATCAGCGGTGGAGACTCCGTTGCAGACGGTCTGATTGTGACACCTGGCATGATAATCCC
GATTACGTAATCTACGGTCTTGATCCATTCGATACTTACCCGAGTCGACTCAGTGTCCAA
CCGTCAGATCTAAACCGAGTCATTGACACGTCGAGTCCGCTTCTCCGCCGACCTTGTGG
CCTCCTTCTTCGCCATTATCGATTCTCCTCCGCTTACTCATGAGTCACCAACCAAGAAGAT
CCAGAGACTAACGACTCCGAAGACGATGACTTCGACCTAGAACCACCTCTCCTCAAAGCT
ATATACGACTGTGCACGGATCTCAGACTCTGACCTAACGAAGCTTCCAAGACGCTTCTT
CAGATCCGAGAATCTGTATCGGAGCTAGGTGATCCGACGGAGCGAGTTGCATTTTACTTC
ACGGAAGCTCTCTCCAACAGACTGTCTCCTAATTGCGCGGCGAGTCGTCTTCTTCTCA
TCTACGGAGGATTTAATCTTATCTTATAAAACCCTAAACGACGCTTGTCTTACTCCAAA
TTCGCACATTTGACGGCGAATCAAGCGATTCTAGAAGCGACGGAGAAGTCGAACAAGATT
CACATCGTCGATTTTGAATCGTTCAAGGTATACAATGGCCTGCTCTTCTTCAAGCTCTA
GCTACTCGTACTTCTGGTAAACCCACTCAAATCCGGGTCTCGGGTATACCCGCTCCATCT
CTCGGTGAATCTCCGGAACCGTCGTTAATCGCCACCGGAAACCGCTCCGTGATTTCCGC
AAGTTCTGGATCTGAATTTGATTTCAATCCCAATTCTCACTCCCATAATTTACTTAAC
GGGTCAAGTTTCCGGGTTCGACCCGATGAAGTACTGCGCGTGAATTTTCATGCTCCAGCTC
TACAAATTACTCGACGAGACGCCGACGATAGTTGACACCGCACTACGGCTCGCCAAATCG
TTGAACCCGAGGGTCGTCACTCTCGGAGAATACGAAGTGAGCTTAAACCGGTTCGGTTTC
GCTAACCGGGTAAAGAACGCGCTTCAATTCATTCGCGGTTTTTCGAATCCCTTGAACCG

AACTTGGGGCGTGATTTCGAGAGAGAGAGTGAGAGTTGAGCGAGAGTTGTTTCGGCCGGAGA
ATCTCGGGTTTGATTGGACCGGAGAAAACCGGAATTCATAGAGAAAGAAATGGAAGAGAAA
GAGCAATGGCGGGTATTAAATGGAGAATGCCGGTTTTGAATCGGTTAAGCTGAGTAATTAC
GCAGTGAGCCAAGCGAAGATTCTATTGTGGAATTACAATTACAGCAATTTGTATTCAATT
GTTGAATCTAAGCCTGGCTTCATCTCTTTGGCCTGGAACGATTACCTCTCCTCACTCTT
TCTTCTGGCGATAA

>G312 Amino Acid Sequence (domain in AA coordinates: 320-336)
MAYMCTDSGNLMAIAQQVQKQKQQEQQQHHQDHQIFGINPLSLNPWPNTSLGFGLSG
SAFPDPFQVTGGGDSNDPGFPFNNLDHHAATTTGGGFRLSDFGGGTGGGEFESDEWMETL
ISGGDSVADGPDCTWHDNPDYVIYGPDPFDTPSRLSVQPSDLNRVIDTSSPLPPPTLW
PPSSPLSIPPLTHESPTKEDPETNDSEDDDFDLEPLLKAIYDCARISDSDPNEASKTLL
QIRESVSELGDPTEVAFYFTEALSNNRSPNSPATSSSSSSSTEDLILSYKTLNDACPYSK
FAHLTANQAILEATEKSNKIHIIVDFGIVQGIQWPALLQALATRTSGKPTQIRVSGIPAPS
LGESPEPSLIATGNRLRDFAKVLDLNFDFIPILTPIHLLNGSSFRVDPDEVLA VNFMLQL
YKLLDETPTIVDTALRLAKSLNPRVVTLLGEYEVSLNRVGFANRVKNALQFYSAVFESLEP
NLGRDSEERVRVERELFGRRISGLIGPEKTGIHRERMEEKEQWRVLMENAGFESVKLSNY
AVSQAKILLWNYNYSNLYSIVESKPGFISLAWNDLPLLTSSWR*

>G1444 (192..1001)

AATCCCTATCCTTCGCAAGACCCTTCTCTATATAAGGAAGTTCATTTCAATTTTTTTTT
GACACGCTGACAAGCTGACTCTAGCATATCTGGCACC GGCGACAGTCCCTTCTTTGGTGC
AAAGATCCCAAAAATCAAATCGAAAGAGAGAATAAATCAAAGGAAGAATCTTTATCT
GCTTTCTCTCGATGAGGATCCGGAACGACAAGTGCCTCTTCTTTATCGTCTCTATTAC
CAGTTCTCTTATCAGATCTCTACTTTAACCGCTCACCGACGGCCACCGCGAGATACTTTC
GCGGTGGTTATAAAGACGGCGGTGATGATTTTGGTTCTCTTCAGCTTTCGCTTCGCCGC
CGTCGCAGATTTCTGATCGGCTTATCAAAGAGATTTGATAAAGAAGAAGGAGGAGGTCA
AGGCTTTGGATGATGATAATGGTGATGTAGACGTCAGAGTCGTACTGATGCATCGGGCA
GCAAGAATGTTAATCCCGGAGGAGAATCCGTCTCTTCAATACAAGTTGTGCGAGAAGAATG
AAAAGGTTGTCTTTGAGGAAGAGAAGAGGCTTTATCAACTTTGAGGATTACGAAGATG
AGGAAGATGAAGAAGCTAGTGGCGGTGGAGGCCGTATTAATAAAGGGAAAAAGAAAGCGA
AAAAGAGCGGTGGTGGGTTAGAGGAAGGATCACGGTGCAGCCGTGTTAACGGTAGAGGAT
GGAGATGTTGTGCAACAACGCTTGTGTTGTTATTCTCTTTGTGAGCATCATCTCGGTAAAG
GAAGGGTAAGGAGCATGAACAAGAGTGGTGGTGGTCGTGGCGGCGAGAAAAAGCGGTGG
TGGTGGAAAGTGAAGAAGAAGAGAGTAAAGCTTGGCATGGTAAAGGCACGTTCAATAAGTA
GTTTGTCTTGACAAACCAGCACTAGTGGTGGTACTAGTGGTGATGTTGATCAGGGTGAGA
TAAGTGCACCTGCTGATCAGTTGCTGTCATGTGATAAGTAGGTCTGTTGATCAGCATTG
CATGTATATGGATATGTGTATGTTTATGTACATGATGATAATGGGCATAGCGGCGCGCT
CTAGACAGGCCTGGAACCGGATCCTCTAGCTAGAGCTTTCGTTAGTATCATCGGGTTTAG
ACAACGTT

>G1444 Amino Acid Sequence (domain in AA coordinates: 168-193)

MRIRKQVPLPLSSLLPVPLSDLYFNRSPTATARYFRGGYKDGDDFGSLQLSLPPPSQI
SDRLIQRLDIKKKEEVKALDDNDGVDVKSRTDASGSKNVNPRGESVSSIQVVEKNEKVV
SLRKRRGFINFEDYEDEEEDDEASGGGGRINKGKKKAKKSGGGLLEEGSRCSRVNNGRWCC
QQTIVGYSLEHHLGKGRVRS MNKSGGGRGGEKKAVVVEVKKRVKLG MVKARSISLLG
QTSTSGGTS GDVDQGEISAPADQFAACDK*

>G801 (27..746)

GATAGTGATAACGAAATCCTAATTCATGCGCGACAACGACGGAGCAGTGAGTAACGGCA
TCATAGTCGAGCAGACGTCAAACAAAGGACCTCTTAACCGCGTTAAGAAACCACCGTCTA
AAGATCGACACAGCAAGTTGACGGAAGAGGAAGAAGGATTTCGTATGCCAATCATTTGCG
CAGCTCGAGTTTTTCAATTGACCAGAGAGTTAGGTCACAAGTCCGATGGTCAAACCATAG
AGTGGCTTCTCCGTCAAGCTGAGCCTTCTATCATAGCCGCCACTGGAAC TGGCACTACTC
CGGCGAGTTTTCTCACTGCTTCTCTCTCACTTCTTCTCCGTTTACTCTCGGGAAACGTG
TCGTGAGAGCGGAGGAAGGAGAAATCCGGCGGCGGAGGAGGAGGGTTAACAGTGGGAC
ACACAATGGGGACTTCGTTAATGGGTGGTGGTGGTCTGGTGGGTTTTGGGCTGTTCGG
CGAGGCCGATTTCGGACAAGCTTGAGCTTTGCAACCGAGCTCCACCGGAAATGGTTT
TTGCGCAGCAGCAGCAACAGCTACACTCTTCGTCCGCCACCAGCAGCAACAGCAAGCTT
CCGCCGCCGAGCAGCTGCAATGGGTGAGGCTTCAGCAGCTAGAGTTGGGAATTATCTTC
CGGGTCATCATCTCAATTTGCTTGCTTCTTTGTCTGGTGGAGCTAACGGGTGGGTGCGA

GGGAAGACGACCACGAACCACGTTGAGAAATGGTATTGTCTTTTTGGTAATGTATAGAAA
AATTCCTATGTTTTATGTCATCGAAAGTGTTTAGAAAGTACCTCTAATTTGCGGTTTCCTT
TTGCTCCTTTTTTACTTAATTTAAGCTTATTGCTTGTGATTAGGGTTTTAGGGTTTAA
GAATATTTGGTCTCGTTAATTTGTTTCGGAGAGTGATAGAAAGAGAGAGATTGATTGA
TTGTTGTACCTAAAACGCTATAAAAGCTCTGTTTTTACTAGCGAAAAAA

>G801 Amino Acid Sequence (domain in AA coordinates: 32-93)
MADNDGAVSNGIIVEQTSNKGPLNAVKKPPSKDRHSKVDGRGRRIIRMPIICAARVFQLTR
ELGHKSDGQTIEWLLRQAEPSSIIAATGTGTTTASFSSTASLSTSSPFTLGKRVVRAEAGES
GGGGGGGLTVGHTMGTSLMGGGSGGFWAVPARPDFGQVWSFATGAPPEMVFAQQQQPAT
LFVRHQQQQASAAAAAAMGEASAARVGNLYLPGHHLNLLASLSGGANGSGRREDDHEPR*
>G1950 (42..764)

CTGAATTCGAACTTTGGAAGAAGAAGCTTTGATCAATCATGGAATTCGAACCGATA
CAGCAAAGCAGATGAGAGACGAAGAGTTGTTCAAAGCAGCGGAATGGGGAGATTCATCGT
TGTTTCATGTCATTATCTGAAGAACAGCTCTCTAAATCTCTCAATTTAGAAACGAAGATG
GTCGCTCTCTCCTCCATGTCGCTGCTTCCCTTCGGCCATTTCTCAAATAGTGAAGTTGTTAT
CAAGTTTCAGATGAAGCAAAGACTGTAATCAATAGCAAGGATGATGAAGGATGGGCTCCTT
TGCATTCCGCTGCTAGCATCGGTAATGCTGAGCTCGTTGAGGTGCTTTTGACCAGAGGTG
CTGATGTCAATGCCAAAAATAACGGTGGTTCGCTGCTCTTCACTATGCTGCTAGCAAAG
GCCGGTTGGAGATTGCTCAGCTTTTATTAAACACACGGTGCAAAGATTAAACATCACAGACA
AGGTTGGTTGCACTCCGCTTCACAGGGCAGCAAGCGTGGGAAAGTTAGAAGTTTGTGAAT
TTCTTATTGAAGAAGGAGCAGAGATCGATGCTACGGATAAAATGGGTCAAATGCACTCA
TGCATTTCAGTTATCTGCGATGACAAACAGGTTGCGTTTCTGCTTATAAGACATGGTGCAG
ATGTGGATGTAGAAGACAAGGAAGGCTACACTGTTCTAGGCCGAGCTACCAATGAATTCC
GACCTGCACTTATCGATGCTGCTAAGGCCATGCTTGAAGGATAAAATGACTCTGGATTAC
TTTAAACTTACTAACTCTGAGAGTTGTTTAGTTACTTAAAGGATTTTCTTTACTGTA
TCATGTTTGCAAATGTTTCTGCCTTATCAATTCATGTTCTGT

>G1950 Amino Acid Sequence (domain in AA coordinates: 65-228)
MEIATDTAKMRDEELFKAAEWGDSSLFMSLSEELSKSLNFRNEDGRSLHVAASFGHS
QIVKLLSSSDEAKTVINSKDEGWAPLHSAASIGNAELVEVLLTRGADVNAKNNGGRTAL
HYAASKGRLEIAQLLLTHGAKINITDKVGCTPLHRAASVGKLEVCEFLIEGAEIDATDK
MQQTALMHSVICDDKQVAFLLIRHGADVDEDKEGYTVLGRATNEFRPALIDAAMKAMLEG
*

>G958 (55..1950)
CGTCGACATGTTTATATTTGTTTCTAGCTAAGAAGTTTGTATAAGGCAGTGGACATGGCT
CCTGTTTTCAATGCCTCCAGGTTTCCGGTTTCATCCAACAGACGAAGAGCTTGTCATATAC
TACCTCAAGCGAAAGATTAATGGTTCGGACTATTGAGTTAGAGATAATACCCGAGATTGAT
CTTTACAAATGCGAACCTTGGGATTTACCTGGGAAGTCCTTGCTGCCAAGTAAAGACCTA
GAATGGTTCTTTTTCAGTCTCTGAGACCGGAAATATCCAACCGGATCAAGAACAACCCGG
GCGACCAAAGCAGGTTACTGGAAAGCCACCGGGAAGATCGTAAAGTGACTTCACATTCA
CGGATGGTTGGAACAAAGAAAACATTAGTTTATTACCGAGGAAGAGCGCCTCATGGCTCT
CGTACCGATTGGGTCATGCACGAGTACCGTCTTGAAGAACAAGAATGTGACTCTAAATCC
GGTATACAGGATGCCTATGCATTTGTGCGAGTATTTAAGAAGAGTGCTTTAGCCAACAAA
ATTGAAGAACAACACCATGGTACGAAGAAGAACAAAGGAACGACTAATAGTGAACAATCT
ACTTCTAGTACTTGTGTTGTTCTGATGGAATGTATGAAAACCTCGAAAACCTCGGGGTAT
CCAGTCTCACCTGAGACAGGAGGCTTAACCTCAACTCGGTAATAATTGTCGTCGGATATG
GAAACGATAGAGAATAAATGGAGTCAGTTTATGTCGATGACACGTCCTTCAACTTCCCA
CCTCAGTCTCAATATGGAACAATCTCATATCCTCCCTCGAAGGTTGATATAGCGTTAGAG
TGTGCAAGACTACAAATCGTATGTTGCCACCAGTACCACCCTTTACGTAGAAGGTCTC
ACACACAATGAATATTTTGGAAACAATGTAGCTAACGATACAGATGAAATGTTGAGCAAG
ATTATAGCATTGGCTCAAGCCTCATGAGCCACGAAACAGTCTAGACTCATGGGACGGT
GGTTCTGCTTCCGGGAACCTTCCATGGAGACTTTAACTATTCCGGAGAAAAAGTCTCATGC
CTAGAGGCGAACGTGGAGGCTGTAGATATGCAAGAACACCATGTGAATTTTAAAGGAAGAA
AGACTTGTGAAAACCTTGAGATGGGTAGGAGTATCAAGCAAGGAACCTGAAAAGAGCTTC
GTTGAAGAACACTCAACGGTAATTCCTATAGAAGATATTTGGAGATATCATAATGATAAT
CAAGAACAAGAACATCATGATCAAGATGGTATGGACGTTAACAACAACAATGGAGATGTG
GATGATGCTTTTCACTCGAGTTTTCGGAAAACGAACATAACGAGAATCTTTTGGACAAG
AACGATCATGAGACAACGAGTTCTCATGTTTGGAGGTGGTAAAAAAGTTGAGGTTAGC

CATGGATTGTTTGTACAACTCGTCAGGTAACCAACACATTCTTCCAACAGATAGTACCA
TCGCAAACCGTTATAGTTTATATAAATCCGACCGATGGCAATGAGTGTGTCATAGTATG
ACATCAAAAGAGGAGGTTTCATGTCCGTA AAAAGATAAATCCGCGAATCAACGGAGTAAGC
TCAACAGTTCTTGGACAATGGAGAAAATTCGCGCATGTTATTGGCTTCATTCTATGCTT
CTATTGATGCGTTGTGTTTCATCGAGGTAACCTAACA AAAACAGAGGCAGTGAAGTTAC
TCGAGGCAGCCTACGAGAGGAGATTGTAACAATCGGGGAACAATACTCATGATGGAAAAT
GCTGTCTGTGAGAAGAAAATTTGGAAGAAGAAGAAAGAGAAAATATGGTTGACGAACAA
GGTTTTCGGTTTCAAGATAGTTTCGTATTGAAGAAGTTGGGGCTTTCTCTGTCTATCATC
TTAGCTGTTTCTACCATAAGTCTTATTTGAATACTGAGGTTCAATATATCATATATGGCT
TTTCACTTTTCTATTGTACTCCCATTTGCCTAGGTCGTATGC

>G958 Amino Acid Sequence (conserved domain in AA coordinates: 7-156)

MAPVSMPPGFRFHPTDEELVIYYLKRKINGRTIELEIIPEDLYKCEPWDLPKSLPSK
DLEWFFFSPRDRKYPNGSRTNRTKAGYWKATGDKRVTSRSMVGTKKTLVYYRGRAPH
QSRTDWMHEYRLEEQECDKSGIQDAYALCRVFKKSALANKIEEQHHGTTKKNKGTNSE
QSTSSSTCLYSDGMYENLENSGYVSPETGGLTQLGNSSSDMETIENKWSQFMSHDTSPN
FPPQSQYGTISYPPSKVDIALECARLQNRMLPPVPLYVEGLTHNEYFGNNVANDTDEML
SKIIIALAQASHEPRNSLSDWGGSSAGNFHGDFFNYSGEKVSCLEANVEAVDMQEHVNFK
EERLVENLRWVGVSKELEKSFVEEHSTVIPIEDIWRYHNDNQEQEHHDQDGMVNNNG
DVDDAFTLEFSENEHNENLLDKNDHETSSSSCFEVVKVEVSHGLFVTTTRQVTNTFFQOI
VPSQTVIVYINPTDGNECCHSMSTKEEVHVRKINPRINGVSSVTLGQWRKFAHVIGFIP
MLLLMRCVHRGNSNKNRGSEGYSRQPTRGDCNNRGITILMMENAVVRRKIWKKKKEKNMVD
EQGFRFQDSFVLKKLGLSLAILAVSTISLI*

>G1037 (1..1722)

ATGACTGTTGAACAAAATTTAGAAGCTTTGGATCAGTTTCTGTAGGAATGAGAGTTCTT
GCTGTTGATGATGACCAAACCTTGTCTCAAATCCTTGAATCTCTCCTTCGTCAGTCCAA
TACCATGTAACAACGACGAACCAAGCACAAAAGGCTTTAGAGTTATTGAGAGAGAACAAG
AACAAGTTTGATCTGGTTATTAGTGATGTTGACATGCCTGACATGGATGGTTTCAAATC
CTTGAGCTTGTGGTCTTGAAATGGACCTACCTGTCTATAATGTTGTCTGCGCATAGTGAT
CCAAAGTATGTGATGAAGGGAGTTACTCATGGTGCTTGTGATTATCTACTGAAGCCGTT
CGTATTGAGGAGTTGAAGAACATATGGCAACATGTCTGTGAGAAGTAGATTGATAAGAAC
CGTGGGAGTAATAAATGTTGATAGAGAGATGGATCAGGTAATGAAGGTGTTGGGAAT
TCTGATCCGAACAATGGGAAAGGTAATAGAAAACGTAAAGATCAGTATAATGAAGATGAG
GATGAGGATAGAGATGATAATGATGATTCGTGTGCTCAAAGAAGCAACGTGTTGTTGG
ACTGTTGAGTGCATAAGAAATTTGTTGCAGCTGTTAACAATTTGGGATATGAGAAGGCT
ATGCCTAAAAAGATTTTGGATCTGATGAATGTTGAGAAGCTCACTAGAGAAAATGTGGCC
AGTCATCTTCAGAAATTCGCTTTACTTGAAGAGGATCAGTGGTGTGGCTAATCAGCAA
GCTATTATGGCAAACCTCTGAGTTACATTTATGCAAATGAATGGACTTGATGGTTCCAT
CACCGCCCAATCCCTGTTGGATCTGGTCAGTACCATGGTGGGGCTCCTGCAATGAGATCT
TTCCCTCCAAACGGGATTCTTGGCAGACTCAATAGCTCTTCGGGGATCGGTGTCCGCAGC
CTTTCTTCTCCTCTGAGGAATGTTCTTGCAAAACCAGACCGATATCGGAAAGTTTCAC
CATGTCTCATCACTTCTCTTAACCACAGTGATGGAGGAAACATACTTCAAGGGTTGCCA
ATGCCTTTAGAGTTGACACAGCTTCAGACAAACAACAACAAAAGTAGAAACATGAACAGT
AACAAGAGCATTGCTGGGACCTCCATGGCTTTTCTTAGCTTCTCTACGCAACAAAACCTCG
CTCATCAGTGCTCCTAATAACAATGTCGTGGTTCTAGAAGGTCACCCACAAGCAACTCCT
CCAGGCTTCCAGGACACAGATCAATAAACGTTTGGAGCATTGGTCAAATGCTGTATCC
TCTTCGACTCACCTCCTCCCCGGCACATAACAGTAATAGTATCAATCATCAGTTCCGAT
GTCTCTCCATTACCGCATTCTAGACCCGACCCCTTGAATGGAACAATGTGTCATCAAGC
TACTCTATACCATTTCTGTGACTCTGCCAATACATTGAGTTCTCCAGCCTTGGATACAACA
AATCCCCGAGCTTTCTGTAGAAACACGACTTCGATTCAAACACAAATGTGCAACCTGGA
GTCTTTTATGGTCCATCCACGGATGCTATGGCTCTGTTGAGTAGTAGTAACCCGAAAGAA
GGGTTTCGTGCTAGGCCAACAGAAGTTACAGAGTGGTGGATTTCATGGTTGCAGATGCTGGT
TCCTTAGATGATATAGTCAACTCCACGATGAAGCAGGTGTGA

>G1037 Amino Acid Sequence (domain in AA coordinates: 11-134, 200-248)

MTVEQNLEALDQFPVGMRLAVDDQTKLILESLLRHCYHVTTTNQAQKALELLRENK
NKFDLVISDVMPDMDGFKLLELVGLEMDLPVIMLSAHSDPKYVMKGVTHGACDYLLKPV
RIEELKNIWQHVVSRFDKNRGSNNNGDKRDGSGNEGVGNSDPNNGKGNRKRKDQYNEDE
DEDRDDNDDSCAQKKQRVVWTVLHKKFVAAVNQLGYEKAMPKKILDLMNVEKLTRENV

SHLQKFRLLYKRI SGVANQQAIMANSELHFMQMNGLDGFHHRPIPVGSGQYHGGAPAMRS
FPFNGILGRNLNSSSGIGVRSLSPPAGMFLQNQTDIGKFHHVSSLPLNHSDDGNNILQGLP
MPLEFDQLQTNMNSRNMNSNKS IAGTSMAPPSFSTQONSLSIPANNVVLVLEGHPOATP
PGFPGHQINKRLEHWSNAVSSSTHPPPAHNSNSINHQFDVSPPLPHSRPDPLEWNNVSSS
YSIPFCD SANTLSSPALDTTNPRAFCRNTDFDSNTNVQPGVFGPSTDAMALLSSSNPK
GFVVGGQKLQSGGMVADAGSLDDIVNSTMKQV*

>G2065 (33..1124)

AACCACACAAAAACAAAAACATATTGATGGGGATGAAGAAGGTAAAGCTATCTT
TGATAGCTAATGAAAGATCAAGGAAAACATCCTTCATGAAGAGGAAAAACGGGATATTCA
AGAAACTCCACGAGTTGTCAACTCTATGTGGTGTCCAAGCTTGTGCTCTCATCTATAGTC
CATTCATACCGGTTCCAGAGTCATGGCCGTCAAGGGAAGGTGCTAAAAAGGTAGCTTCAA
AGTTTCTGGAGATGCCGCGGACAGCCGAAACCAGGAAGATGATGGATCAAGAAACCCATC
TTATGGAGAGGATTACCAAAGCAAAAGAGCAACTAAAGAATTTGGCTGCTGAGAACCGAG
AATTACAGGTTAGACGATTTATGTTTGTGTTGAAGGCAAAATGTCCAGTATCGTT
ATGATGCAAAAGACCTTCAAGATTTGCTATCTTGTATGAATCTATATCTCGATCAGCTTA
ACGGAAGGATCGAGTCCATTAAAGAAAACGGTGAGTCGTTGTTGCTTCCGTCTCTCCTT
TTCCTACTAGAATTGGTGTGACGAAATTTGGTGATGAGTCGTTTTCCGACTCTCCTATTC
ATTCTACAAC TAGGGTTGTAGATACTCCTAATGCTACCAATCCTCATGTTCTTGCGGGCG
ATATGACTCCTTTCTTGATGCGGACGCAAAATGCGGTAAGTCTCCAGTCGATTTTCTG
ATCATATTCAATATGAAAATATGAATATGAGTCAAAATCTGCATGAACCGTTTCAACACC
TTGTTCTCTACTAACGTTTGTGATTTTATCAAAATCAGAATATGAATCAGGTTCAATACC
AGGCTCCTAATAATCTGTTTAAATCAGATTCAACGAGAATTCTACAACATAAAATTTGAATC
TGAATTTGAATCTGAATTCAAATCAGTATCTGAATCAACAACAATCATTCATGAATCCGA
TGGTGGAACAACATATGAATCATGTTGGAGGGCGTGAAAGCATTCTTTCTGCGGACAGAA
ACTACTACAAC TACAATCAACTACCAGCCGTTGATCTTGCTTCCACCAGTTACATGCCTT
CAACCACCGATGTTATGATCCTTACATCAACAACAATCTCTAATCACAAAAGACGGAGA
TTTTCTAGTTTAA

>G2065 Amino Acid Sequence (domain in AA coordinates: TBD)

MGMKKVKLSLIANERSRKTSFMKRKNGIFKKLHELSTLCGVQACALIYSPFIPVPESWPS
REGAKKVASKFLEMPRTARTTRKMDQETHLMERITKAKEQLKNLAAENRELQVRRFMFDC
VEGKMSQYRYDAKLDQLDLSCMNLVLDQLNGRIESIKENGESLLSSVSPFPTRIGVDEIG
DESFSDSPIHSTTRVVDTPNATNPHVLADMTPLDADANAVTAPSRFSDHIQYENMMS
QNLHEPFOHLVPTNVCDFYQNQNMNQVQYQAPNNLFNQIQREFYNINLNLNLNLSNQYL
NQQQSFMNPMVEQHMHNVGGRESIPFVDRNYNYNQLPAVDLASTSYMPSTTDVYDPYIN
NNL*

>G2137 (77..1123)

GGGATTTGACTTTAGCACTTCAAAATCCAAAGCTAAAAGACAAAAAGAATAGAGGTTCCG
ATTTGCATCTCCATTAAATGGGCATCGATCTTTCTCTTAAGCTCGAGGCCGAGGAGAAAAA
GAAAGAGATAGAAGGATCGAAACATAGCCGTGAGAACAAGAAGACGAAGAACATGATGC
TAGTGGTGATGAAGATGAACAAATGGTGAAAGAAGACGAAGATGATTCTTCTTTAGG
TTTAAAGAACCCGAGAAGAAGAAAACGAACGTGAAGAGCTCTTGCAGCTACAGATCCAGAT
GGAAAGTGTGAAAGAAGAGAATACTAGGTTGAGGAAGCTTGTGAGCAGACTCTTGAAGA
TTATCGTCATCTTGAGATGAAATTCCTCGTTATCGATAAAACCAAGAAGATGGATCTTGA
AATGTTCTTGGAGTACAAGGCAAACGATGTGTGGATATAACAAGTAAGGCTCGGAAAAG
AGGAGCTGAGAGATCTCCGTCAATGGAAAGAGAAATAGGGCTTTCATTTCTCTAGAGAA
AAAAACAGAAACAGAAGAGAGCAAAGAAGCTGTTCAAGTCTCATCACCAAGATACAATAG
TAGCAGCTTAGATATGAATATGCCACGTATCATTTTCATCTTCTCAAGGTAATAGAAAGGC
CAGGGTGTCCGTGAAGGCGAGATGTGAGACCGCAACAATGAATGATGGATGCCAATGGAG
GAAGTACGGTCAGAAAACCGCGAAAGGAATCCATGTCCTCGAGCTTATTACCGATGCAC
CGTGGCTCCAGGATGTCCCGTTAGAAAACAGGTGCAAAGGTGTTTGAAGACATGTCAAT
ACTGATAACAACCTACGAAGGAACACATAACCATCCACTTCCGGTCCGAGCAACAGCCAT
GGCTTCCACTGCCTCTACTTCTCATTTCTTGTTACTCGATTCCAGTGACAACCTCTCTCA
TCCTTCTCTATTACCAAACTCTCAAGCCATAGACTCTTCTTTGATTACATACCCACAAAA
TAGCAGCTACAACCAATCGAACCATAGAAGCTTGAACCTTGATGGTCCATCTAGAGGAGA
TCACGTTTTCATCTTCTCAAAACCGATTAAATTTGGATGATGTAGAGTTTCTATATCTCTA
TGCTTGTCTTTGGTCCCATTATTTGTCAATTATGGATTCTTTGCCTTTCTTCTGTTCTC
GTTTCTAACATTTATGTTTCGTATA

>G2137 Amino Acid Sequence (conserved domain in AA coordinates:109-168)

MGIDLSLKLAEKKKEIEGSKHSRENKEDEEHSDASGDEDEQMVKEDEDDSSSLGLRTRE
EENEREELLQLQIQMESVKEENTRLRLVEQTLQEDYRHEMKFPVIDKTKMDLEMFLGV
QGKRCVDITSKARKRGAERSPSMEREIGLSLSLEKKQKQESKEAVQSHHORYNSSSLDM
NMPRIISSSQNRKARVSVRARCETATMNDGCQWRKYGQKTAKGNPCPRAYYRCTVAPGC
PVRKQVQRCLEDMISILITTYEGTHNHPLPVGATAMASTASTSPFLLLDSSDNLSHPSYYQ
TPQAIDSSSLITYPQNSSYNNRTIRSLNFDGPSRGDHSVSSQNRLNWMW*

>G746 (1..1311)

ATGGGTGAGGAGTTAGCTGACACAATGAACCTGGATTTGAATCTTGGGCCTGGTCCTGAG
TCTGATCTCCAACCTGCACCAAACGAGACTGTGAATTTGGCTGATTGGACTAATGACCCG
CCTGAGAGATCTTCTGAAGCTGTGACAAGGATCAGGACTCGGCATAGGACACGGTTCAGA
CAGCTTAATCTCCCGATCCCGGTTCTATCTGAAACCCATACCATGGCTATAGAGCTCAAC
CAGTTGATGGGAAATTTCTGTAAATAGAGCTGCTATGCAGACTGGTGAGGGTAGTGAAAGA
GGCAATGAGGATTTGAAAATGTGTGAGAATGGCGATGGAGCCCTTGGGGACGGTGTATTG
GATAAGAAAAGCGGATGTCGAGAAAAGCAGTGGCAGCGACGGTAACCTTTTCGATTGTAAT
ATATGTTTGGATTTGTGCAAGGAGCCGGTTCTCACCTGTTGTGGTCATCTTACTGTTGG
CCTTGTCTGTACCAATGGTTACAAATTTCCGATGCAAAGGAATGTCCTGTTGTAAAGGA
GAGGTGACCTCCAAACCGTGACACCGATCTATGGACGTGAAAACCAAGAGAGAAATT
GAAGAGAGTTTAGATACTAAGGTCCCCATGAGACCACACGCGAGACGCATTGAGAGCTTG
AGGAATACAATTCAAAGGTCCGCTTTTACAATACCAATGGAAGAAATGATTAGACGTATA
CAGAATAGGTTTGACAGGGATTCAACCCAGTCCCTGATTTTAGTAACCGAGAGGCATCA
GAAAGAGTCAACGATCGAGCCAATTTCGATCCTTAACCGGTTGATGACATCTAGGGGAGTT
AGATCAGAGCAGAACCCAGGCTAGTGTGTCAGCAGCAGCCATTGTGCGAGCATCAGAGGAT
ATTGATCTAAATCCAAACATTGCTCCTGATCTTGAAGGAGAAAGCAACACGAGATTCCAT
CCTCTCTTGATCAGGAGACAGTTACAGTCGCACCGAGTTGCAAGGATCTCGACTTTCCT
TCTGCGTTGAGTTTCAGCTGAGAGGCTTGTGGATGCGTATTTTAGGACTCATCCGTGGGG
AGGAACCACCAAGAGCAAAACCATCATGCTCCTGTTGTGGTTGATGATAGAGACTCATTC
TCAAGCATTGCAGCTGTTATAAACTCTGAGAGTCAAGTGGATACTGCAGTTGAGATCGAT
TCTATGGCTCTTTTCGACATCGTCTCGAGGAGAAGGAATGAGAATGGTTTCGAGGGTTTCT
GATGTAGACAGTGCAGATTCTCGTCCGCTTAGGAGAAGGAGATTTACTTGA

>G746 Amino Acid Sequence (domain in AA coordinates: 139-178)

MGEELADTMNLDLNLGPGPESDLQAPAPNETVNLADWTNDPPERSSEAVTRIRTRHTRFR
QLNLPPIVLPVLTSETHTMAIELNQLMGNSVNRAAMQTGEGSERGNEDLKMCENGDGLDGV
DKKADVEKSSGSDGNFDFCNICLDLSKEPVLTCCHLYCWPCLYQWLQISDAKECPVCKG
EVTSTKVTPIYGRGNHREIEESLDTKVPMPHARRIESLRNTIQRSPFTIPMEEMIRRI
QNRFRDSTPVPDFSNREASERVNDRANSILNRLMTSRGVRSEQNQASAAAAAIVAASED
IDLNPNIAPDLEGESNTRFHPLLRQLQSHRVARISTFTSALSSAERLVDAYFRTHPLG
RNHQEQNHHPVVDVDRDSFSSIAAVINSESQVDTAVEIDSMALSTSSRRRNENGSRV
DVSADSRPPRRRRFT*

>G2701 (46..837)

GTGTTTGTAGTTGAACTTATTCTTCCCTTTTGTGTTTTTAGGTATGGAGACTCTGCAT
CCATTCTCTCACCTACCTATCTCTGACCACCGGTTTCGTTGTTCAAGAGATGGTGAGCTTA
CACAGCTCGAGTAGCGGTAGCTGGACTAAAGAAGAGAACAGATGTTTCAACGAGCTCTT
GCGATATACGCTGAAGACTCGCCTGATCGCTGGTTTAAAGTTGCTTCCATGATCCCTGGA
AAGACTGTTTTGTATGTTATGAAGCAATATAGTAAGCTTGAAGAAGACGTTTTTCGATATT
GAAGCAGGACGTGTTCCCATTCCTGGTTATCCTGCAGCTTCTTCCCTTGGGGTTTGAC
ACCGACATGTGTGCTAAACCGCCTAGTGGAGCTAGAGGATCTGATCAAGATCGAAAGAAA
GGAGTCCCTTGGACAGAGGAAGAACACAGGAGATTCTTGTTAGGCCCTTCTCAAGTACGGT
AAAGGAGATTGGAGAAACATATCGAGAACTTCGTGGTGTCAAAGACGCCAACGCAAGTG
GCGAGCCACGCCCCAAAGTATTACAGAGACAGCTCTCCGGAGCCAAGGACAAACGCAGG
CCAAGTATCCATGACATCAACCCGGCAATCTTCTCAATGCCAATCTCAACCGTTCCCTT
TCCGATCATAGAGATATTCTCCCTGATTTAGGGTTTATCGATAAGGATGATACGGAGGAG
GGAGTAATATTTATGGGTGAGAATCTCTCTTCAGAAAATCTGTTTTCTCCATACCAACT
TCATTGGAAGCTGCCATTAACTTCGCCGAGAAAATGTCTTCAGTGCCCGGAGCTTAAGGC
AACATAGAATCCCCAAACTCAGCGGC

>G2701 Amino Acid Sequence (domain in AA coordinates: 33-81, 129-183)

METLHPFSLPISDHRFVQEMVSLHSSSSGWSWKENKMFERLAIYAEDSPDRWFKVA

SMIPGKTVFDVMKQYSKLEEDVFDIEAGRVPIPGYPAASSPLGFDTDMCRKRPSGARGSD
QDRKKGVPTWTEEEHRRFLGLLKYKGDWNRNISRNFVVSKTPTQVASHAQKYYQRQLSGA
KDKRRPSIHDIITGNLLNANLNRSFSDHRDILPDLGFDKDDTEEGVIFMGQNLSENLF
SPSPTSFEAAINFAGENVFSAGA*

>G1819 (1..639)

ATGGAAGAGAACACGGAACAACAACCACTACCTGCCGCAACCATCGTCTTCCCAACTG
CCGCCGCCACCATTGTATTATCAATCAATGCCGTTGCCGTCATATTCACTGCCGTCGCCG
TACTACCGCAGATGCGGAATTATTGGATTGCGCAGATGGGAAACGCAACTGATGTTAAG
CATCATGCGTTTCCACTAACCAGGATAAAGAAAATCATGAAGTCCAACCCGGAAGTGAAC
ATGGTCACTGCAGAGGCTCCGGTCCCTTATATCGAAGGCCTGTGAGATGCTCATTCTTGAT
CTCACAATGCGATCGTGGCTTCATACCGTGGAGGGCGGTCCGCAAACTCTCAAGAGATCC
GATACGCTCACGAGATCCGATATCTCCGCCGCAACGACTCGTAGTTTCAAATTTACCTTC
CTTGGCGACGTTGTGCCAAGAGACCCCTCCGTCGTTACCGATGATCCCGTGCTACATCCG
GACGGTGAAGTACTTCCCTCCGGGAACGGTGATAGGATATCCGGTGTGTTGATTGTAATGGT
GTGTACGCGTCACCGCCACAGATGCAGGAGTGGCCGGCGGTGCCTGGTGACGGAGAGGAG
GCAGCTGGGGAAATTGGAGGAAGCAGCGGCGTAATTGA

>G1819 Amino Acid Sequence (domain in AA coordinates: 46-188)

MEENNGNNNNHYLPQSSSQLPPPLYQSMPLPSYSLPLPYSPQMRNYWIAQMGNATDVK
HHAFPLTRIKKIMKSNPEVNMVTAEPVLISKACEMLILDLTMRSLHTVEGGRQTLKRS
DTLTRSDISAATRSFKFTFLGDVVPDRPSVVTDDPVLHPDGEVLPPTVIGYPVFDENG
VYASPPQMGEWPAVPGDGEEAAGEIGSSSGN*

>G1227 (372..1451)

TCTTCCGTGTGTTAACAGAAGTCCCCACAATTGTCTGTCTTCGCTGCGAGACAAAACCTGC
CACAGCCAATAATGTTTCTCTGAGGGACCTTGCTTCTGTCTCAGAGACTCGCTCTCTCTCTC
CTCTTCTTGCTCTGCTCAGCTCTCTCACCAACTCATCTTCAGTCCTCAAACAAACATCTG
TTCTCATCTTTGTTTCTTCTTCTTCTTCTCATATCTCATTTTCAATTTTCCCAATTTT
TCTTCAACATCTTCATAGCAATTTAAGACCACTATTCCATTATAAAGCTAACTGCTTTAG
AAACTCCTCACATTTATTTCTTCCCCATCATTTGTTTTAGAGAGGGAGAAAGAAAAAGAGC
TCAGCTTTCTGATGGAGAGGAGTATTCAAGGACAAAACAAGCTCTGTTGTTTGGACAAA
AAGTGAATGTGAGAAGAAGCCTACAAGTTCAAGAACTGTAGAGGATCATCAAAGCTTTG
CCCTTGAAGAGGAAGAACAACAACCTCTCAACTCCGAGCTTGCTGCAAGACACAACAATAC
CATTTCTACAAATGCTGCAACAAAGTGAAGACCCCTTACCCTTTTGTCTATTCAAAGACC
CAAGCTTTCTAGCACTACTATCTCTCCAGACACTTGAAAAGCCTTGGGAACTCGAAAAC
ACCTCCACATGAAGTTCCAGAGTTTCATTACCGATCCATTCTGAAACCAACCACTACT
ATCATAATCCATCTTTGGAAGGAGTCAATGAAGCCATCTCAAACCAAGAACTTCCATTCA
ACCCACTAGAGAATGCGCGTTCAAGACGCAAGCGGAAAAACAACAACCTTGGCATCATTTGA
TGACAAGAGAAAAAGCGAAGAGAGAAGAACTAAACCAACAAGAACATAGAAGAGATAG
AGAGTCAAAGAATGACACACATTGCGGTTGAACGAAACCGCAGACGCCAAATGAACGTTT
ATCTGAACCTCACTCCGCTCCATCATTTCCATCTTCATACATCCAGAGGGGAGACCAAGCGT
CAATAGTAGGAGGAGCAATAGACTTCGTAAAGATCCTAGAGCAACAGTTGCAATCCCTTG
AAGCACAAAAGAGAAGTCAACAGAGTGATGATAACAAAGAGCAAATTCCAGAAGATAACA
GTCTCAGGAACATTTTCGTGCAACAAGTTGCGTGCGAGTAATAAAGAAGAACAAAGTAGCA
AACTCAAAATCGAAGCCACAGTGATAGAGAGTCACGTCAACCTAAAAATTCAATGTACGA
GGAAACAAGGACAACTTCTCAGATCAATCATATTGCTGGAGAACTTCGATTCACTGTTC
TTCATCTCAACATCACATCTCCGACCAATACATCTGTCTCTTATTCCTTCAACCTCAAGA
TGGAAGATGAATGTAATTTGGGATCAGCGGATGAGATAACGGCGGCGATTTCGTCAGATTT
TCGACAGCTGATTGACTAATCCAAGTAAAAAGTAAAAAAGAAACGTTTACTTTG
GTAACCTTCGTTTTCATGATTAAATTTCTTTATTTGGTCGTATGTGATTGGAGTCTTCTCGG
CATGGAACCTTGACTTTGGTTTTAGGGTACTAGTCTCTACAGAAGCTGTGGTCCTTCTTTG
GATGC

>G1227 Amino Acid Sequence (domain in AA coordinates: 183-244)

MERSIQGNKLCCLDQKVNVRSLQVQETVEDHQSFALIEEEEQQLSTPSLLQDTTIPFLQ
MLQQSEDPSPFLSFKDPFLALLSLQTLKPEWLENYLPHEVPEFHSPHSETNHYHNP
SLEGVNEAISNQELFPNPLENARSRRKRKNNNLASLMTREKRKRRTKPTKNIEEIESQR
MTHIAVERNRRRQMNVLNLSLRSIIPSSYIQRGDQASIVGGAIDFVKILEQQQLQSLEAQK
RSQQSDDNKEQIPEDNSLRNISSNKLRSNKEEQSSKLKIEATVIESHVNLIKIQCTRKQG
QLLRSIILLEKLRFITVLHLNITSPTNTSVSYSFNLKMEDECNLGSADEITAAIRQIFDS*

>G2417 (118..1311)

CATACCGGTGGAAGATTCTGCTTTACTACGCTCTCCGCTTCTTCTTCTCCTCGATTTCGAT
TCTCCTCATGGGTTTATCATGAATTTTAGGTTTTAGTAATTTCAGAACTCGAGTGATG
ATCCCGAATGATGATGATGATGCAAAATCTATGAAGAATTATCCGTTAAATGATGATGAT
GCAAAATCTATGAAGAATTATCCGTTAAATGATGATGATGCAAAATCTATGGAGAATTAT
CCGTTAAGGTCAATTCCGACGGAGCTTTCACACACTTGTTTCATTGATACCACCTTCTTTA
CCAAACCCTTCAGAAGCAGCAGCAGACATGTCCTTCAATTCAGAACTCAATCAAATCATG
GCAAGGCCTTGTGATATGCTCCCTGCCAATGGTGGAGCTGTTGGTCATAACCCTTTTTTG
GAACCAGGATTCAACTGCCCCGAGACAACAGATTGGATTCCCTCTCCACTCCCCCATATT
TATTTTCTTCGGGTTCTCCCAATCTAATAATGGAGGATGGTGTCATTGATGAGATTAC
AAACAAAGTGACTTGCCACTTTGGTATGACGACTTGATTACCACTGATGAAGATCCACTC
ATGTCTAGTATCTTGGGCGATCTTCTCTTGACACTAATTTCAACTCAGCTTCAAAGGTG
CAGCAACCAAGTATGCAATCGCAGATTCAACAACCCCAAGCTGTTCTGCAGCAGCCTTCT
TCTTGTGTGGAATTGCGCCACTTGATAGGACAGTATCCTCAAACAGCAACAACAATAGC
AACAGTAATAATGCAGCAGCAGCAGCTAAGGGACGTATGCGTTGGACGCCTGAACCTCAT
GAGGTTTTTGTGACGCTGTTAACCAGCTCGGTGGCAGTAATGAAGCAACTCCTAAAGGT
GTCCTGAAGCATATGAAAGTCGAAGGTTGACTATTTTTCATGTCAAAAGTCATTTCGAG
AAATATAGAACAGCTAAATATATACCAGTACCATCAGAAGGTTCCGCCGAGGCAAGGTTG
ACACCGCTTGAGCAAATTACATCTGATGATACGAAACGTGGGATAGATATCACTGAGACT
CTGCGAATTCAGATGGAACATCAGAAGAACTGCATGAGCAGCTTGAGAGTCTAAGAACA
ATGCAACTTCGGATAGAAGAGCAAGGAAAGGCGCTGTTGATGATGATTGAGAAGCAAAAT
ATGGGTTTCGGCGGACCAGAACAAGGAGAGAAAACAAGTGCAGAAAACGCCTGAAAATGGT
TCAGAGGAGTCGGAATCCCCGCGGCCAAAGCGTCCGAGAAATGAAGAATGAAGGAAACCT
TTCTTCGGATGTAGATCATAAACTGTGGTTTTGGTGGAGTTGTAGAGTATGACTTATT
AGGAGTAGAGCTTTCAGTCTTCTTCAGGC

>G2417 Amino Acid Sequence (domain in AA coordinates: 235-285)

MIPNDDDDANSMKNYPLNDDDDANSMKNYPLNDDDDANSMENYPLRSIPTELSHTCSLIPPS
LPNPSEAAADMSFNSELNQIMARPCDMLPANGGAVGHNPFLBPGFNCPETTDWIPSPLP
IYFPGSPNLIMEDGVIDEIHKQSDLPWYDDLITTEDPLMSSILGDLDDLTFNFSASK
VQPSMQSQIQQPQAVLQQPSSCVELRPLDRTVSSNSNNNSNNNAAAAAKGRMRWTPEL
HEVFVDVAVNQLGGSNEATPKGV LKHMKEGLTIFHVKSHLQKYRTAKYIPVPEGSPEAR
LTPLEQITSDDTKRGIDITETLRIQMEHQKLHEQLESRLTMQLRIEEQKALMMIEKQ
NMFGFGPEQGEKTSAKTPENGSESESPRPKRPRNEE*

>G2116 (104..1117)

TTTCATCTCCATCATTTATCTCCATTGACATTGTTCTCAATTGCGAATAATAATCATAATTA
TTCACACAACCAAGCATCTCTCAGATTCTCTTAAAAAATGGAGAAATCAGATCC
TCCACAGTCCCAAAGCCCGCGCCACTATTATCCCTCCTCCGATCCAATTCCTAATGC
CGATCCGATTCCATCTTCTCTCTCCACCGCGATCTCGCTCCGACGATATGTCCATGTT
CATGTTTCATGGATCCCCTCTCTCCCGCCGACCACTTCTCCGACGACCTTCCCTCCGA
CGACGATCTCTTCTCTCTTTCATCGATGTCGATAGCCTCACCTCTAATCCCAATCCCTT
TCAAAATCCTTCCCTCTCTCTCAACTCCGTTTCCGGCGCTGCTAATCTCTCTCTCTCC
TTCTCTCTCGTCTCGCCACCGTCACAGCAATTCCGTTGACGCTGGATGCGCCATGTATGC
CGGTGATATCATGGACGCTAAGAAAGCTATGCCTCTGAAAACTCTCTGAGCTTTGGAA
CATCGATCCCAAACGCGCCAAAAGGATTCTAGCGAATCGACAATCTGCAGCTCGATCCAA
AGAGAGAAAAGCTCGATACATTCAAGAACTTGAGCGCAAAGTTCAATCTCTTCAAACCGA
AGCTACCACTCTCTCTGCTCAGCTTACTCTCTACCAGAGAGACACAAATGGACTAGCAAA
CGAAAACACAGAGCTGAAACTTAGGTTGCAAGCAATGGAACAACAAGCTCAGCTTCGTAA
TGCTTTTAAACGAAGGTTGAGGAAAGAGTTGAAAGGATGAAGATGGAGACAGGAGAAAT
CTCTGGTAATTGAGATTGTTTTGATATGGGAATGCAGCAGATTGAGTATTCTTCTCAAC
TTTCATGGCTATTCCACCATATCATGGCTCAATGAACCTCCATGATATGCAGATGCATTG
TAGTTTCAATCTATGGAGATGTCCAATCTCTCAAAGCGTGTGCGACTTTCTACAGAACGG
CCGAATGCAAGGCTGGAGATTAGTAGCAATAGCTCAAGCTTAGTCAAATCTGAAGGACC
TTCTCTCTCTGCTAGTGAGAGTAGCTCTGCCTATTGACGACAAGATTATGATGAGGCTCA
TTTTTCTG

>G2116 Amino Acid Sequence (conserved domain in AA coordinates:150-210)

MEKSDPPFPVKPATIIPSSDPIPNADPIPSSSFHRRSRSDMSFMFMDPLSSAAPPSS
DDLPSDDDLFSFIDVDSLTSNPNPFQNPNSLSSNSVSGAANPPPPSSRPRLRHSNSVDA

GCAMYAGDIMDAKKAMPPEKLSELWNIDPKRAKRILANRQSAARSKERKARYIQELERKV
QSLQTEATTLSAQLTLYQRDTNGLANENTELKLRLQAMEQQAQLRNALNEALRKEVERMK
METGEISGNSDSFDMGMMQIQYSSSTFMAIPYPYHGSMNLHDMQMHS SFNPMEMSNSQSVS
DFLQNGRMQGLEISSNSSSLVKSEGPSLSASESSAY*

>G647 (1..948)

ATGATGATCGGCGAAAATAAAAAACGGGCCACATCCAACGATCCATATCCCTCAATGGGAT
CAAATCAACGATCCAACGGCCACAATCTCTTCACCATTCCTCTCCGTCAACCTTAACAGC
GTTAACGACTACCCACACTCTCCGTCACCGTATCTCGACTCCTTCGCTTCTCTCTTCGGT
TACCTCCCCTCAAACGAGTTAAACAAACGATTTCAGACTCATCAAGTGGCGACGAGTCATCA
CCACTCACCGACTCATTTCTCTCCGACGAGTTTGCATCTACGAGTTCAAATCCGGCGCA
TGCGCTCGAGGTGATCTCATGATTGACGGAGTGTCGTTCCGCATCTCCGGAGAAAA
GCTCGACGACGTGATCCGAGAAAGTTTATTACTCCGGCACCGCTTGCTCGAGTTTCGT
AAAGGAAGTTGTAGAAGAGGTGATTCTGTGTGAGTTCTCTCATGGAGTTTTCGAGTTTGG
CTCCATCCTTCTCGTTACCGTACTCAGCCGTGTAAAGACGGAACTAGCTGCCGGAGAAGA
ATCTGTTTTCTTCGCTCATACGACGGAGCAGTTACGTGTATTACCTTGTTCTGTAGATCCA
GATCTTGGATTCTTCTCAGGATTAGCTACTTCTCCGACTTCGATTCTTGTTTCTCCTTCG
TTTTACCACCGTCGGAATCTCCGCCGCTTTCTCCGAGTACCGGTGAACCTTATTGCGTCG
ATGAGGAAAAATGCAATTGAACGGAGGTGGTTGTTCTGTGGAGTTCTCCGATGAGATCTGCA
GTTAGGTTACCTTTTTCTCGTCGTCTCTGCGTCCGATTACGGCGGCAACGTGGCCGAGGATA
GTAGAGTTTGTAGATCGAAGAGCTCCGGCATGGAATTTGTGGAATCTGGGAAAGAGCTG
AGAGCGGAGATGTATGCAAGACTCAGTAGAGAGAACCTACTCGGTTGA

>G647 Amino Acid Sequence (domain in aa coordinates: 77-192)

MMIGENKNRPHPTIHIPOWDQINDPTATISSPFSSVNLNSVNDYPHSPSPYLDSSFASLFR
YLPSNELTNDSDSSSGDESSPLTDSFSSDEFRIYEFKIRRCARGRSHDWTECPFAHPGEK
ARRRDPRKFHYSGTACPEFRKGCRRGDSCEFSHG VFECWLHPSRYRTQPCDKGTSCRRR
ICFFAHTTEQLRVLPCSLDPDLGFFSGLATSPTSILVSPSFSPPPSESPLSPSTGELIAS
MRKMQLNGGGCSWSSPMRSVRLPFSSSLRPIQAATWPRIREFEBIEAPAMEFVESGKEL
RAEMYARLSRENSLG*

>G974 (377..1162)

AAAAAAAAAAGTTGATATACTTTCTGGTTTCTCCTTAAC'TTTTATTCTTTACAAATCCAT
 CCCCCTTAGATCTGTTTATTTCCCGCTACTTTGATTCA'TTTCTGTTAGTAATCTGTCTTT
 CGTATAGAAGAAA'CTGATTTCTTGGTTTGTATTTTCTTAAAGAGATCAATCTTTT
 TTTT'TGATCTTCTTGTGTTT'TTTTTTCTTTGTAGAATTAATCGTTTGTGAGGGTATTTTT
 TTAATTCCCTCCTCTCAGAAATCTACACAGAGGTTT'TTTTATTTATAAACCTCTTTTTCG
 ATTTTCTTGAAAA'CAAAAACTCTGTTCTTACTTTT'TTTTACAAGAACAGGGAAAAAA
 TTTCTTTTATTAGAAA'AGACAAC'TTCTATGAGATTTTACAGTAACAAACGTTTCAACA
 ACTGTGATCCATT'CGGTGGTGAATTAATGAAGCGCTTTTACCTTTTATCAAAAGCCCTTC
 CAACGATTCA'TCCGCGTTTGCCTTCTCTACCCGCTCCAATTCATACGGGTCGGATCT
 CCACTCATTTTCTCACCATCTTAGTCTTAAACCGGTCTCAATGAAACAAACCGGTACTTC
 CGCGGCTAAACCGACGAAGCTATACAGAGGAGTGAGACAACGCTCACTGGGGAAATGGGT
 GGCTGAGATT'CGTTTACCGAGGAATCGAACTCGACTTTGGCTCGGAACATTTCACACGGC
 GGAGGAAGCTGCTTTAGCTTATGACAAGGCGGCGTATAAGCTCCGAGGAGATTTTGCGCG
 GCTTAATTTCCTGATCTCCGTATAACGACGAGTATCAACCTCTTCAATCATCAGTCGA
 CGCTAAGCTTGAAGCTATTTGTCAAAACTTAGCTGAGACGACGCAAGAACAGGTGAGATC
 AACGAAGAAGTCTTCTTCTCGGAACGTTTCACTCAACCGTCGCAGTGAAACTACCGGAGGA
 GGACTACTCTAGCGCCGGATCTTCGCCGCTGTTAACGGAGAGTTATGGATCTGGTGATC
 TTCTTCGCCGTTGTTCGGAGCTGACGTTTGGTGATACGGAGGAGGAGATTACGCCCGTG
 GAACGAGAACCGCTTGGAGAAGTATCCGTCGTACGAGATCGATTGGGATTGATTCCTTCA
 GTGTTTCGAGTCTTGTAAATTAGATGTTGCCATAGGGGTATTTTAGGGACTTTAGAGCTCT
 CTGCGATGGAGT'TTTTGGTCATTGCAGAGATTTTATTATTATTAAGGGGGTTTGTATGT
 TAATATCAAATAAGTTTATCTACTTTTGATGTTAATTAGTGTAACTCTCTCGCTCGGTCCA
 AGCTGTTTTTTTTTGGCATGCTTCGACCGTGTGAGATTTCTTATGTAATTTTGTAGTTC
 CTTGATTTTTCTTAGTTCAAGTTAAATTTGGCACAAAAA'AAAAAAAAAAAAAAAA

>G974 Amino Acid Sequence (domain in AA coordinates: 81-140)

MTTSMDFYSNKTFFQSDPFGGELMEALLPFIKSPSNDSSAFASLPAPISYGSDLHFSFH
HLSPKPVSMKQTGTSAAKPTKLYRGVRQRHWGKWVAEIRLPRNRTLWLGTFTAEAAAL
AYDKAAAYKLRGDFEARLNFPDLRHNDYEOPILOSSVDAKLEAICONLAETTOKOVRSTKKSS

SRKRSSTVAVKLPEDDYSSAGSSPLLTESYSGSGSSSPLSELTFGDTEEEIQPPWNENAL
EKYPSYEIDWDSILQCSSLVN*

>G1419 (27..692)

GAAGACTCCAACATAATTCATCATCTATGGCTTCTTCACATCAACAACAGCAAGAACAAG
ACAGTCAGCTTTAGATCTCATAACCCAACACCTTCTTACTGATTTCCCTTCTTAGACA
CCTTTGCCTCCACATCCACCACTGCACCACCTCAACTCTAAGCCAACGCAAACCACTC
TTGCCACTATAGCAGTTCTTACTACTGCACCGGTGGTTCAAGAGAATGATCAAAGGCATT
ACAGAGGCGTCAGGAGAAGACCATGGGGTAAGTATGCGGCTGAGATCAGAGACCCAAACA
AGAAAGGTGTTTCGTGTCTGGTTAGGCACTTTGTACACAGCCATGGAAGCTGCAAGAGGTT
ATGACAAGGCAGCTTTTAAACTACGAGGAAGCAAAGCTATTCTTAACTTCCCACTTGAAG
CAGGAAAGCATGAGGACTTGGGAGACAACAAGAAGACTATTTCTTTAAAGCAAAGAGGA
AGAGACAGGTGACGGAGGATGAAAGCCAGCTGATCAGCCGTAAAGCTGTTAAGAGGGAAG
AAGCTCAGGTTACAGCTGATGCTTGTCCATTAAACGCCATCAAGTTGGAAGGGGTTTGGG
ACGGAGCAGACAGTAAAGACATGGGAATATTTTCCGTGCCCTCTGTTATCTCCTTGTCCAT
CTCTTGGACACTCTCAACTCGTAGTTACTTAAAGCTTCAGAGGGTCAAACGGAAAAAATC
AACATTGGATTGTTTTCAAAGCTTCTAGATTAGCTGATTGTAAAAAATGTTTTACTATA
TTCATTCACTTCTTCTTAAATGCAATTCTTCTACCCCTCC

>G1419 Amino Acid Sequence (domain in AA coordinates: 69-137)

MASSHQQQEQDQSSALDLITQHLLTDFPSLDTFSTLHCTTSTLSQRKPPLATIAVPTT
APVVQENDQRHYRGVRRRPWGKYAAEIRDPNKKGVRVWLGTFTAMEAARGYDKAAFKL
GSKAILNFPLEAGKHEDLDGNKKTISLKAQRKRQVTEDESQILSRKAVKREEAQVQADAC
PLTPSSWKGFWDGADSKDMGIFSVPLLSFPCPSLGHSQLVVT*

>G1634 (22..855)

TTATCTCGTAGCCTTTAAACGATGGAGACTCTGCATCCACTACTCTCGCACGTGCCAACT
TCTGACCACCGTTTGTAGTTCAAGAGATGATGTGCTTGCAAAGCTCGAGCTGGACTAAA
GAAGAGAACAAGAAGTTTGAGCGAGCTCTTGCTGTCTACGCTGATGACACGCCGTGATCGC
TGGTTCAAAGTTGCTGCTATGATCCCTGGAAAGACCATATCAGATGTATGAGGCAATAC
TCTAAGCTTGAAGAAGACCTCTTCGATATCGAAGCAGGACTTGTCCCGATCCCGGTTAC
CGTTCAGTTACTCTTGTGGATTGATCAGGTTGTGAGTCCACGTGACTTTGATGCGTAT
CGTAAACTTCTTAATGGAGCCAGAGGATTTGATCAAGACCGTAGGAAAGGAGTTCCATGG
ACGGAGGAAGAACAACAGGAGATTCTTGTTAGGGCTTCTCAAGTATGGGAAAGGAGATTGG
AGAAACATATCGAGGAACCTTTGTGGGATCAAAAACACCAACTCAGGTTGCAAGTCATGCC
CAAAGTACTACCAAAGACAGCTTTCCGGTGCGAAAGACAAACGACGGCCTAGCATTCAC
GACATCACCACCGTCAATCTTCTCAATGCCAATCTTAGCCGTCCATCGTCTGATCACGGT
TGCTTAGTCTCAAAACAGGCCGAGCCGAAACTAGGGTTACCCGACAGGGATAATGCAGAG
GAGGGAGTTATGTTCTTGGTCAGAATCTATCCTCGGTCCTTCTTCTTCTACGATCCTGCC
ATTAAGTTTTCCGAGCAAATGTTTACGGTGAAGGAGGTTACTGTATCTCACAAGATCTT
GAAACGAGAAAAATGAGAATTTTGAATTTTAACTATTGCAACGAAACATAATTGC

>G1634 Amino Acid Sequence (domain in AA coordinates: 129-180)

METLHPLLSHVPTSDHRFVVQEMMCLQSSSWTKEENKKFERALAVYADDTDRWFKVAAM
IPGKTISDVMRQYSKLEEDLFDIEAGLVPIPGYRSVTPCGFDQVVSPRDFDAYRKLPNGA
RGFDQDRRKGVPTWTEEHRRFLLGLLKYKGDWRNISRNFVGSKTPTQVASHAQKYYQRQ
LSGAKDKRRPSIHDIITVNLNANLSRPSSDHGCLVSKQAEPKLGFTDRDNAEEGVMFLG
QNLSSVFSSYDPAIKFSGANVYEGGYCISQDLETRK*

>G1637 (1..954)

ATGGTGGAAGGAGACGGTGACGGTGGCGAAAAACGTGCTCACACTGTGGCCATAATGGCCAT
AACGCACGGACTTGTCTCAACGGCGTTAATAAGGCAAGTGTTAAACTGTTCCGGCGTTAAT
ATATCGTCTGATCGGATTAGGCCGCTGAGGTAACGGCGTTAAGGAAGAGTCTTAGTTTG
GGAAACCTTGATGCTCTTCTCGCTAACGATGAAAGTAACGGTAGCGGTGATCCTATCGCC
GCCGTTGATGATACCGGTTATCATTCCGATGGTCAGATTCAATCCAAGAAGGGTAAAACT
GCTCATGAGAAGAAAAAGGGGAAGCCATGGACGGAAGAAGACATCGTAATTTCTTAATC
GGTTTAAACAACTCGGAAAGGAGATTGGAGAGGCATTGCAAAGAGTTTCGTGTCGACA
AGAACACCAACACAAGTCGCAAGTCATGCTCAGAAATATTTTATTAGGTTAAACGTTAAC
GACAAAGAGAAAAAGACGTGCTAGTCTCTTGACATCTCTCTCGAAGATCAGAAGGAGAAA
GAGAGGAACCTCTCAAGATGCTTCAACAAAGACTCCACCTAAACAACCAATAACCGGAATT
CAACAACCGGTAGTACAAGGTCTACTCAAAACGAGATTTCGAACAGGTTTCAGAATTTA
TCAATGGAGTATATGCCAATCTACCAACCCATACCACCTTACTACAACCTTCCACCTATT

ATGTACCATCCAAATTATCCAATGTACTATGCCAACCCCTCAAGTACCGGTTAGGTTTGTTCATCCTTCTGGTATACCTGTTCCAAGACATATACCGATTGGTTTGCCTCTGTCTCAACCGAGTGAAGCTTCTAATATGACAAATAAAGACGGTTTGGATCTTCATATCGGTTTGCCTCCA CAAGCTACTGGAGCTTCTGACTTGACTGGTCATGGCGTTATTCATGTGAAATGA

>G1637 Amino Acid Sequence (domain in AA coordinates: 109-173)
MVKETVTVAKTCSHCGHNGHNARTCLNGVNKASVKLFVGNISSDPIRPPEVTALRKSLSLG NLDALLANDESNNGSDPIAAVDDTGYHSDGQIHSKKGKTAHEKKKGKWPTEEHNRNFLI GLNKLKGKDWRGIAKSFVSTRPTQVASHAQKYFIRLNVDNKRKRASLFDISLEDQKEK ERNSQDASTKTPPKQPITGIQPPVVGHTQTEISNRFQNLMEYMPIYQPIPPYNYFPPI MYHPNYPMYANPQVPVRFVHPSGIPVPRHIPIGLPLSQPSEASNMTNKGDLDLHIGLPP QATGASDLTGHGVIHVK*

>G1818 (601..1161)
TAACAAATCAAATAATTAGAGAAATAACCAAAATTTAACTTTTAGAGGGACTACAGGATT TGTACTTTGTACATTCATATATTATGTATATATCGTTTCATACATTAATTTGAACCAA TGTAATTAAGTAAATTCATTTAACATCATGAGCAAATCTTATTAATTAATCTCTTAA AATTTTGAGCAAATATGCTTTTACATTTAACATTTGAAAACATCATTTTAAACAAGATA TTCAAAACTAAGTTTGTACAGCAAATTTTAACTTTCAATTTTATAGAGAAAAAGGTAT TTTTTTTTTTGTTCATTTTATAAGACTATTATTTGGTATATAATATACACTTTAAGTA AAAACAAATCTCTTTCTTTTCTTCTTATAATACCAACCACAAGTCTGTCAGTCACACA CACACAGTTAATAACATTAAATATCTTAACAACTACTAAATAGGTTGAGATTCATATA TGTAAAGAGATCACTTCTTAATCTTATCCTACCATATCTTATATACGCTTAATTTTCCTT TATATATGCAAACTCCACATAAAAAATATCTCAAACCCAAACACTTCAAACAAAAAATA ATGGAGAACAACAACAACAACCACCAACAGCCACCGAAAGATAACGAGCAACTAAAGAGT TTCTGGTCAAAGGGGATGGAAGGTGACTTGAATGTCAAGAATCACGAGTCCCCATCTCT CGTATCAAGAGGATAATGAAGTTTGATCCGGATGTGAGTATGATCGCTGCTGAGGCTCCA AATCTCTTATCTAAGGCTTGTGAAATGTTTGTTCATGGACCTCACGATGCGTTCATGGCTC CATGCTCAAGAGAGCAACCGACTCACGATACGGAATCTGATGTTGATGCCGTAGTGTCTT CAAACCGTCATCTTTGATTTCTTGGCTGATGATGTCCTAAGGACGAGGAGAGCCCGTT GTCGCCGCTGCTGATCCTGTGGACGATGTTGCTGATCATGTGGCTGTGCCAGATCTTAAC AATGAAGAACTGCCGCCGGGAACGGTGATAGGAATCCGGTTTGTACGGTTTAGGAATA CACGCCCCACACCCGAGATGCCCTGGAGCTTGGACCGAGGAGGATGCCACTGGGGCAAT GGAGGAAACGGTGGAATTAATATTGGATTGGGTTTTGTAACCGCTGTTGTGAGAACTT GAATTTCTTTTGTAGTTCTGCTTATGTTTTCAATGTTATGTTTTTTAGTTGTGAATGTA TTTCTGTTGTTTTGTCCAAAAAAGAAATGTATTCTGTTGTTGTCTTTCAAATGA ATCTAATGGTTTATGAATATTGGCTTTAGATTAATTTATGCATACAAAAACACAAGGATT ACGGATAAAAAAGTCCTCAGTTTACCCATGGAACATAATCTTCTAGTGATTCCTTATGA GAGTAGAAAAAGAAATCATATATTATAATCTATTTTATAAGAGATAGGGTACTGTAAACAAG GATGTTTATTCGGCTATTTCTTTTAAATCACTTTTACTTGTCAAGACTCTTTTGT GTTTGCAGCTTTTGTGATTACATTCTAGAGGCAACAAGATCCAGAGATCTAGCAAAA AAAACTTATTTGAAACCTGAATCTATTTTAAAAATTTTCCAACCTCATTTTTCGTTCTTA TTCTTTGTTTTCCAACGGAATTTGGCGCAAAACGATTATTTGAATTTGTCTTTCAAG

>G1818 Amino Acid Sequence (domain in AA coordinates: 36-113)
MENNNNNHQQPPKDNEQLKSFWSKMEGDLNVKNHEFPISRIKRIMKFDPDVSMIAAEAP NLLSKACEMFVMDLTMRSWLHAQESNRLTIRKSDVDVVSQTVIFDFLRDDVPKDEGEPPV VAAADPVDDVADHVAVVDLNNNEELPPGTVIGTPVCYGLGIHAPHPQMPGAWTEEDATGAN GNGNGN*

>G1820 (1..609)
ATGGCTGAGAACAAEAACAACAACGGCGACAACATGAACAACGACAACCACCAGCAACCA CCGTTCGTACTCGCAGCTGCCGCCGATGGCATCATCCAACCCCTCAGTTACGTAATTACTGG ATTGAGCAGATGGAACCGTCTCGGATTTCAAAAACCGTCAGCTTCCATTGGCTCGAATT AAGAAGATCATGAAGGCTGATCCAGATGTGCACATGGTCTCCGCAGAGGCTCCGATCATC TTCGCAAAGGCTTGCGAAATGTTTCATCGTTGATCTCAGGATCGGTCGTTGGCTCAAAGCC GAGGAGAACAAACGCCACACGCTTCAGAAATCGGATATCTCCAACGCAGTGGCTAGCTCT TTCACCTACGATTTCTTCTGATGTTGTCCCTAAGGACGAGTCTATCGCCACCGCTGAT CCTGGCTTTGTGGCTATGCCACATCCTGACGGTGGAGGAGTACCGCAATATTATTATCCA CCGGGAGTGGTGATGGGAATCTTATGTTGGTAGTGGAATGTACGCGCCATCGCAGGCG TGGCCAGCAGCGGCTGGTGACGGGGAGGATGATGCTGAGGATAATGGAGGAAACGGCGGC

GGAAATTGA

>G1820 Amino Acid Sequence (domain in AA coordinates: 70-133)

MAENNNNNNGDNMNDNHQQPPSYSQLPPMASSNPQLRNYWIEQMETVSDFKNRQLPLARI
KKIMKADPDVHMVSAEAPIIFAKACEMFIVDLTMRSLKAEENKRHTLQKSDISNAVASS
FTYDFLLDVVPKDESIATADPGFVAMPHPDGGGVPPQYYPPGVVMGTPMVGSGMYAPSQA
WPAAAGDGEDDAEDNNGNGGGN*

>G1903 (1..1200)

ATGTCTAAATCTAGAGATACGGAGATAAAGTTGTTTGGGAGGACAATCACATCTCTTTTA
GATGTGAATTGTTATGATCCGTCGTCGTTGTCCCCTGTTACAGATGTTTCTTCTGATCCA
AGCAAGGAGGATTTCGTCTTCTTCTTCATCTTCTTGTCTCCAACCTATTGGACCAATCAGG
GTTCCGGTTAAAAAAGTGAGCAAGAGAGTAACAAATTCAAAGATCCATATATATTATCC
GATCTAAACGAACCAACCAAGCAGTATCTGAGATTTCATCACCAAGAAGTTCCAAGAAC
AACTGTGATCAACAGAGCGAGATCACAACAACCACTACCACAAGTACTACATCAGGAGAG
AAATCAACGGCTCTCAAGAAACCGGACAAGCTTATTCCATGTCCTAGATGTGAAAGCGCA
AACACCAAATCTGTTATTACAACAACCTACAACGTGAACAGCCACGTTACTTCTGCAGG
AACTGTGAGAGGTATTGGACAGCTGGTGGATCTATGAGGAACGTTCCCTGTTGGCTCAGGT
CGTCGCAAGAACAAAGGATGGCCTTCTTCAAACCATTAATTGCAAGTCACTTCTGAGGAT
TGTGATAATAATAACTCGGGGACGATCCTTAGTTTCGGTTCCTTCGGAGTCTTCGGTTACA
GAGACTGGTAAGCATCAGTCAGGTGATACAGCAAAGATAAGTGCTGATTGAGTTTCTCAA
GAAAAATAAAGCTACCAAGGTTCTTCTCCGCAAGTAATGTTACCTAATAATTCTTCT
CCTTGGCCTTACCAATGGAGTCCAACGGGTCCTAACGCTAGTTTCTACCCTGTCCCCTTC
TACTGGGGATGCACGGTTCGGATATACCTACCTCAGAGACTTCATCATGTTTAGGAAAA
CGGTCAAGAGATCAAACCTGAAGGAAGAATCAATGATACTAATAACAATAACTACTACA
AGAGCAAGATTGGTCTCAGAATCTCTTAGAATGAATATCGAAGCTAGTAAGAGCGCTGTG
TGGTCTAAGTTACCGACAAAACCCGAGAAAAAACGCAAGGATTGAGTTGTTCAATGGA
TTTGACACAAAGGGAAACAGCAACAGAAGTAGCTTGGTCTCCGAACTTCTCACAGTCTA
CAAGCAAACCTGCAGCGATGTCTAGAGCTATGAACCTCAGGGAGAGCATGCAACAATAA
>G1903 Amino Acid Sequence (domain in AA coordinates: 134-180)
MSKSRDTEIKLFGRTITSLLDVNCYDPSSLSPVHDVSSDPSKEDSSSSSSSCSPTIGPIR
VPVKKSEQESNKFDPYILSDLNEPPKAVSEISSPRSSKNNCDQQSEITTTTTTSTTSGE
KSTALKKPKDLIPCPCRESANTKFCYNNYNVNQPRYFCRNCQRYWTAGGSMRNPVVGSG
RRKNKGWPPSSNHYLQVTSEDNDNNSGTTILSFGSSESVTETGKHQSGDTAKISADSVSQ
ENKSYQGLFPPQVMLPNNSSWPYQWSPTGPNASFYPVPFYWGCTVPIYPTSETSSCLGK
RSRDQTEGRINDTNTTITTTTRARLVSESLRMNIEASKSAVWSKLPKPEKKTQGFSLFNG
FDTKGNNSNRSSLVSETSHSLQANPAAMSRAMNFRESMQQ*

>G371 (1..582)

ATGGAGATTGAGAAGGATGAGGACGACACAACATTGGTTGATTCTGGAGGAGACTTCGAC
TGCAACATATGTTTGGATCAGGTTTCGAGACCCGGTCGTGACTTTATGTGGCCACCTGTTT
TGTTGGCCCTGCATTACAAAGTGGAATTATGCGTCCAACAATTCAAGACAACGAGTCGAT
CAATACGATCATAAGAGGGAAACCAACAAATGTCCGGTATGCAAATCTGATGTCTCCGAG
GCTACGCTTGTCCCCTGATCTACGACGAGGACAGAAAGCTCCCGAGTCCGGTTCAAATGTA
CCGAGCAGACCAACTGGTCCGGTTTATGACTTAAGAGGAGTTGGTCAACGTTTAGGAGAA
GGGAGAGTCAACGTTACATGTATAGAATGCCGTGATCCGGTGTGGTGTGGTATGCGAA
ATGGTATACCGGAGACTATTTGGAGAGTCTTCGAGCAACATGGCACCTTACCGCGATATG
AATGTCCGGTCTAGGCGACGGGCAATGCAGGCTGAGGAGTCATTAAGCAGAGTCTACTTG
TTTCTACTTTGCTTCATGTTTATGTGTCTATTCTCTTCTAA

>G371 Amino Acid Sequence (domain in aa coordinates: 21-74)

MEIEKDEDDTTLVDSGGDFDCNICLDQVRDPVVTLCGHLFCWPCIHKWTYASNNRQRVD
QYDHKREPPKCPVKSDVSEATLVPIYGRGQKAPQSGSNVPSRPTGPVYDLRGVQRLGE
GESQRYMYRMPDPVMGVVCEMVYRRLFGESSNMAPYRDMNVRSSRRRAMQAEESLSRVYL
FLLCFMFMCLFLF*

>G597 (255..1310)

AAAATTCTCCTGTAAAATTAAATATTATAAAAGTGGTTTCTTTTTCATTTATGTTTATAT
AATTTTCATCTTTAATCTTAAATTCTGGTAACCTTAATGCGCGATCCGCTTTTCTAAAGT
TTTGTGAGAGAGAAGAGATCTAAAAAATCCACAATTTTGTTCAAATCTTGAGTTAAAT
GCTGAATTTTAGGCCTTGTGCTTAGATTTATGGCTTAAAGTTTCAAACCTTTTCATTGGA
TATGTGAGAAGAAAATGTCAGGATCTGAGACGGGTTTAAATGGCGCGACAGAGAATCAA

TGCAATTTACAATGGCTCTCCACCAGCAGCAGCAACACAGTCAAGCTCAACCTCAGCAGT
CTCAGAACAGGCCATTGTCAATTCGGTGGAGACGACGGAACCTGCTCTTTACAAGCAGCCGA
TGAGATCAGTATCACCACCGCAGCAGTACCAACCCAACCTCAGCTGGTGAGAATTCTGTCT
TGAACATGAACTTGCCCGGAGGTGAGTCTGGAGGCATGACTGGAAGTGAAGTGAGCCAG
TGAAGAAAGAGAGAGGTAGACCGAGGAAATATGGGCCTGATAGTGGTGAAATGTCACCTTG
GTTTGAATCCTGGAGCTCCTTCTTTCACTGTCAGCCAACCTAGTAGCGGCGGCGATGGAG
GAGAGAAGAAGAGAGGAAGACCTCCTGGTTCTTCTAGCAAAAGGCTCAAGCTTCAAGCTT
TAGGCTCGACTGGAATCGGATTTACGCCCTCATGTACTTACCGTGCTGGCTGGAGAGGATG
TATCATCCAAGATAATGGCGTTAACTCATAATGGACCCCGTGCTGTGTGTGTCTTGTCTG
CAAATGGAGCCATCTCCAATGTGACTCTCCGCCAGTCTGCCACATCCGGTGGAACTGTTA
CATATGAGGGGAGATTTGAGATTCTGTCTTTATCGGGATCTTTCCATTTGCTGGAGAACA
ATGGTCAAAGAAGCAGGACGGGAGGTCTAAGCGTGTCATTATCAAGTCCGGATGGTAATG
TCCTCGGTGGCAGTGTAGCTGGTCTTCTTATAGCAGCATCACCTGTTGAGATTGTTGTTG
GGAGTTTCTTACCAGACGAGGAGAAAAGAACCAGCAATGTTGGGACAAATGGGACTGT
CGTCACCCGTATTACCGCGTGTTGGCCCCAACGCAGGTGCTGATGACTCCAAGTAGCCAC
AATCTCGAGGCACAATGAGTGAGTCATCTTGTGGAGGAGGACATGGAAGCCCTATTTCATC
AGAGCACTGGAGGACCTTACAATAACACCATTAAACATGCCCTGGAAGTAGCCAAGTGATC
TGTGTGCGGCTTAAACCAACAACCTTCCCGTTATTAGAGTGATTTATTTCTACATTTGGTT
TAGACTTTCTAGTTCTGATGGTTATTTCTACAGTTGGTTTAGACTTTCTAGTTCTGTTCA
GACAAAAGGAGTTTGATAAATTGACCGACCTATTTTGTGTGTTTGAGGTACTTTTCAGAAC
CATAGGTGTTTCAGAAATTAGAATGTTCTGTTTAAAAA
>G597 Amino Acid Sequence (domain in AA coordinates: 97-104,137-144)
MSGSETGLMAATRESMQFTMALHQQQHSQAQPQQSQNRPLSFGDDGTALYKQPMRSVS
PPQQYQPNsAGENSVLNMLPGGESGGMTGTGSEPVKKRRGRPRKYGPDSEGMSLGLNPG
APSFVTSQPSGGDGGKEKRRPPGSSSKRLKLQALGSTGIGFTPHVLTVLAGEDVSSKI
MALTHNGPRAVCVLSANGAISNVTLRQSATSGGTVTYEGRFEILSLSGSFHLLNNGQRS
RTGGLSVSLSSPDGNVLGGSVAGLLIAASPVQIVVGSFLPDGEKEPKQHVGMGLSSPVL
PRVAPTQVLMTPSSPQSRGTMSSESSCGGHGSPHQSSTGGPYNNNTINMPWK*
>G1009 (28..1704)
AAAAAAAAAAAAACCTATTCCCAAAGATGAAGAACAATAACAACAAATCTTCTTCTTCT
TCTAGCTATGATTCTTCTTTGTCTCCTTCTTCTTCATCCTCCTCCCACCAGAACTGGCTC
TCTTTCTCTCTCTCCAACAATAACAACAACCTCAATTCTTCTCCTCAAACCTAATCTCACT
TCCTCCACATCAGATCATCATCATCCTCACCCTTCTCACCTCTCTCTCTTCAAGCTTTC
TCCACTTCTCCAGTCGAACGGCAAGATGGGTACCGGGAGTTTCACCCAGCGATGCCACG
GCGGTTCTTTCCGTATACCCCGCGGTCTTAACTTGAGAACTTCTTCGGCGGAGGAGCC
TCAACGACGACAACAAGACCAATGCAACAAGTGCAATCTCTTGGCGGCGTGTCTTCTCT
TCCGACCTACAGCCACCGCTTCATCCTCCGTCCGCCGCGAGATCTACGACTCTGAGCTC
AAGTCAATAGCCGCTAGCTTCTTAGGAACTACTCCGGTGGACACTCGTCCGAGGTCTCT
AGCGTACATAAACAACAACCGAATCCTCTAGCTGTCTCAGAGGCTTCGCCTACTCCGAAG
AAGAACGTAGAGAGTTTTGGACAACGTACCTCGATTATAGAGGAGTCACAAGACATAGA
TGGACTGGAAGATACGAAGCTCATCTATGGGATAATAGTTGCCGAAGAGAAGGCCAAAGC
AGAAAAGGAAGACAAGTTTATTTAGGTGGTTATGATAAGGAAGATAAAGCAGCTAGAGCT
TACGACCTTGCGACTCTTAAGTATTGGGGTCTTACAACCTACGACTAATTTCCCGATATCA
AATTACGAATCTGAACTTGAAGAAATGAAACACATGACTCGACAAGAGTTCTGTGCTTCT
TTAAGACGGAAAAGCAGTGGATTCTCTAGGGGTGCCTCCATGTACAGAGGCGTCACTAGA
CATCATCAGCATGGTCGATGGCAGGCACGAATTGGAAGAGTTGCAGGCAACAAGACCTT
TATCTTGGCACATTTAGCACTCAAGAGGAAGCTGCAGAAGCTTATGATATAGCAGCGATC
AAATTCGCGCGTCTAATGCAGTCACCAATTTGACATCAGTCGATATGATGTCAAATCA
ATTGCTAGCTGTAATCTCCCTGTGGGTGGACTAATGCCTAAACCTTCTCCAGCAACCGCA
GCGGCTGACAAAACCGTTGATCTTCTCCATCCGACTCTCCATCTTAACCACACCGTCC
CTCACGTTCAATGTGGCAACACCGGTCAATGACCATGGAGGAACCTTTTACCACACTGGT
ATACCAATCAAACAGACCCGGCTGATCATTATTGGTCCAACATCTTTGGATTCCAGGCA
AACCCGAAAGCAGAAATGCGACCATAGCAAACTTTGGGTCCGATCTTCATAACCTTCT
CCTGGTTATGCTATAATGCCGTAATGCAGGAAGGTGAAAACAACCTTTGGTGGTAGTTTT
GTTGGGTCTGATGGGTATAACAATCATTTCCGCTGCATCGAACCCGGTCTCAGCAATTCCG
CTGTCTCGACAACATAATGAGTAACGGTAACGAAGGGTATGGTGGAACATAAACTGG
ATTAATAACAACATTTCAAGTTCTTACCAAACCTGCAAAATCAAATCTCTCTGTTTGCAC

ACACCGGTTTTTGGGTTGGAATGAGTATTCACATCTTAGTGAGAACTAAAATAAATATGT
AGGAAAAAATAAGGCTCTGTTTGAAGAAATCAGATATTTCTTCTTAGATTATTTAAGT
AGTTTAAAAAATATTTTTAAGTGTTTCACTTTTACGTTTGTCTGCTGACCACGAATT
TTGCTGGATCTGACAGTACTAATCTTTGTTTAAATGACCTTATGGGTTCCCTTTTTTACTT
TCCAGAACTTTTATTTACTTTTTTCTTCATTTTTTCTTCATTTTTTTTGTGTGGGACAAT
ATGAATGATTGAAGATGGAACTGCTTGCATGTGAATAACGAAATCAAACNATCTTCG
GTAACCTAAAAA

>G1009 Amino Acid Sequence (domain in aa coordinates: 201-277, 303-371)

MKNNNNKSSSSSYDSSLSPSSSSSSHQNWLSFSLSNMNNNFNSSSNPNLTSSTSDHHHP
HPSHLSLFQAFSTSPVERQDGSPPVSPSDATAVLSVYPGGPKLENFLGGGASTTTTRPMQ
QVQSLGGVVFSSDLQPLHPPSAEIIYDSELKSIASFLGNYSGGHSSEVSSVHKQPNP
LAVSEASPTPKNVESFGQRTSIYRGVTRHRWTGRYEAHLWDNSCRREGQSRKGRQVYLG
GYDKEDKAARAYDLAALKYWGPTTTTNFPI SNYSELEEMKHMTRQEFVASLRKSSGFS
RGASMYRGVTRHHQHGRWQARIGRVAGNKDLYLGTFTSTQEEAAEAYDIAAIKFRGLNAV
NFDISRYDVKSIA SCNLPVGGMLPKPSPATAAADKTVDLSPSDSPSLTTPSLTFNVATPV
NDHGGTFYHTGIPKDPADHYWSNIFGFQANPKAEMRPLANFGSDLHNPSPGYAIMPVM
QEGENNFSGSFVGS DGYNNHSAASNPFVSAIPLSSTTTMSNGNEGYGGNINWINNNISSY
QTAKSNLSVLHTPVFGLE*

>G170 (1..1107)

ATGGGGATGAAGAAGGTGAAGCTATCTTTGATAGCTAATGAAAGATCAAGGAAAACATCC
TTCATAAAGAGGAAAGACGGGATTTTAAAGAACTCCACGAGTTGTCAACTCTGTGTGGT
GTCCAAGCTTGTGCTCTCATCTACAGTCCATTACATACCGGTTCCAGAGTCATGGCCGTCA
AGGGAAGGTGCTAAAAGGTGGCTTCAAGGTTTCTGGAGATGCCGCCGACAGCCCGAACC
AAGAAGATGATGGATCAAGAGACTTACCTTATGGAGAGGATTACCAAAGCAAAAGAGCAA
CTAAAGAACCTGGCTGCTGAGAACCGAGAGTTACAGGTTAGACGATTTATGTTTGATTGT
GTTGAAGGCAAAATGTCCCAGTATCATTATGATGCAAAAGACCTTCAAGATTTGCAATCT
TGTATAAATCTATATCTCGATCAGCTTAACGGAAGGATCGAGTCCATTAAAGAAAATGGT
GAGTCGTTGTGTCTTCCGTCTCTCCTTTTCTACTAGAAATGGTGTGTGACGAAATGGT
GATGAGTCATTTTCCGACTCTCCTATTCATGCTACAACCTGGGGTTGTAGATACTCTTAAT
GCTACCAATCCTCATGTTCTTACGGGCGATATGACTCCTTTTCTTGATGCGGACGCAACT
GCGGTAAC TGCTCCAGTAGATTTTTTGATCATATTCCATATGAAAATATGAATATGAGT
CAAAATCTGCATGAACCGTTTCAACACCTTGTTCTACTAACGTTTGTGATTTTTCAA
AATCAGAATATGAATCAGGTTCAATACCAGGCTCCTAATAATCTGTTTAAATCAGATTCAA
CGAGAATTCTACAACATAAATTTGAATCTGAATTTGAATCTGAATTCGAATCAGTATCTG
AATCAACAACAATCATTATGAATCCGATGGTGGAAACAACATATGAATCATGTTGGAGGG
CGTGAAAGCATTCCTTTCTGTGACGGAACTGCTACAACCTACCATCAACTACCATCCAAT
CAACTACCAGCCGTTGATCATGCTTCCACCAGTTACATGCCTTCCACCACCGGTGTCTAT
GATCCTTACATCAACAATAATCTCTAA

>G170 Amino Acid Sequence (domain in aa coordinates: 2-57)

MGMKKVKLSLIANERSRKTSFIKRKDGIFKKLHELSTLCGVQACALIYSPFIPVPESWPS
REGAKKVASRFLEMPPTARTKKMMDQETYLMERITKAKEQLKNLAAENRELQVRRFMFDC
VEGKMSQYHYDAKDLQDLQSCINLYLDQLNGRIESIKENGESLLSSVSPFPTRIGVDEIG
DESFSDSPIHATTGVVDTLNATNPHVLTGDMTPFLDADATAVTASSRFFDHI PYENMNMS
QNLHEPFQHLVPTNVCDFFQONQNMNQVQYQAPNNLNFNQIQREFYNINLNLNLNSNQYL
NQQQSFMNPMVEQHMHNVGGRESIPFVDGNCYNHQLPSNQLPAVDHASTSYMPSTGVY
DPYINNNL*

>G1768 (185..1426)

CTTCCTTTTGTCTCAGCTGCGAGCTTTGGTTGGATCTCTCACTTGCAAAACCAAATCCCT
TATCGACTTCCACCGAAAGATCACTTCTTAACCTACACAAGGTGTTTGTATGAAGATCA
GATAAATAAAGGTCATTTGAGGATAATGGTTGATGTTCAAAGATTCTTACTTGCTTATT
TGTGATGGACAATGTAAGAGGTTCAATAATGTTGCAGCCACTGCCAGAGATAGCTGAGAG
TATCGATGATGCTATCTGCCATGAACCTCCATGTGGCCTGATGATGCTAAAGATTGTT
ATTGATAGTGGAGCAATATCAAGGGGAGACTTGAAGTTGGTACTTGTGTGCTTGTGCAAA
AGCTGTTTTCTGAGAATAATCTTCTAATGGCACGATGGTGTATGGGTGAGTTGCGCGGTAT
GGTTTCGATTTCTGGTGAGCCAATCCAGAGATTGGGAGCTTATATGTTAGAAGGGCTTGT
TGCTAGGCTTGCTGCTTCTGGTAGTTCGATATATAAGTCTCTCCAGTCCAGAGAACCAGA
GAGTTATGAATTTTATCTTATGTGTATGTTCTGCATGAGGTTTGTCCATATTTCAAGTT

TGGATACATGTCAGCGAATGGTGCATTGTCAGAAGCAATGAAGGATGAAGAGAGGATTCA
CATTATTGACTTCCAAATTGGACAAGGGAGCCAGTGGATAGCACTTATCCAGGCTTTTGC
AGCTAGGCCTGGTGGGGCTCCAAATATTGCAATTACCGAGTTGGTGATGGATCTGTCTT
GGTTACAGTCAAGAAGAGACTAGAGAACTTGCAAAGAAGTTTGATGTTCCATTGAGGTT
CAATGCGGTTTCAAGGCCAAGTTGTGAAGTTGAAGTGGAAAATCTTGATGTCCGAGATGG
CGAAGCCCTTGGAGTGAACCTTGCTTACATGCTGCATCATTTGCCAGATGAGAGTGTAAAG
CATGGAAAACACAGGGACCGGTTGCTGAGGATGGTGAAGAGTCTATCACCTAAAGTAGT
CACTCTTGTTGGAACAAGAATGCAACACGAACACTTCCCCCTTTCCTTCTAGGTTCCCTGA
GACATTAAAGTTATTACACGGCAATGTTTGAATCTATCGATGTTATGCTTCCGAGAAATCA
CAAGGAAAGGATCAATATCGAGCAGCACTGCATGGCAAGGGATGTCTGCAACATCATAGC
TTGTGAAGGAGCCGAGAGGATCGAAAGACACGAGCTTCTCGGGAAATGGAAGTCAAGGTT
TTCCATGGCGGGTTTGGAGCCATACCCCTTGAGCTCAATCATTTCAGCCACCATTAGAGC
CCTCTTGAGAGATTACGCAACCGGATGCGATTGAAGAAAGAGATGGTGCTCTGTACCT
TGGTTGGATGGACCGAATCTTGGTCTCATCTTGTGCATGGAAGTGAAGAATAAACGTCT
CCAAGAATGTAATGCAAAAGACAGAACTGGAAGTAATAGATAGTTTTGTCTCATAACCAT
TAATAAGGTTGAATCAAATCATATACATCCCCATGCTACAACCTATTACACAGGCTCCATC
AACAAAGAAGGGCTCTTGTGTGTACCTTCTTCTCTGTAACCTTATTTTGAACCAAAT
GGAAGTGGTTACAT

>G1768 Amino Acid Sequence (domain in AA coordinates: 54-413)
MDNVRGSI MLQPLPEIAESIDDAICHEL SMWPDADKDLLLIVEAISRGDLKLVLVACAKA
VSENNLLMARWCMGELRGMVSI SGEPIQRLGAYMLEGLVARLAASGSSYKSLQSREPES
YEFLSYVYVLHEVCPYFKFGYMSANGAIAEAMKDEERIHIIDFQIGQGSQWIALIQAFAA
RPGGAPNIRITGVGDGSLVTVKKRLEKLAKKFDVPFRFNAVSRPSCEVEVENLDVRDGE
ALGVNFAYMLHHL PDESVS MENHRDRLRMVKSLS PKVVT LVEQECNTNTSPFLPRFLET
LSYYTAMFESIDVMLPRNHKERNIBQHCMARDVNNIIACEGAERIERHELLGKWKSRFS
MAGFEPYPLSSII SATIRALLRDYSNGYAIERD GALYLGWMDRILVSSCAWK*

>G185 (77..988)
ATGCAAAAATAACATAGTAACAATACTTTAAACTATTTACACCACCTTAACTCTTATTCT
CCACTCTTTGAACGTAATGGAGAAGAACCATAGTAGTGGAGAGTGGGAGAAGATGAAGAA
CGAGATCAACGAGCTAATGATAGAAGGAAGAGACTATGCACACCAGTTTGGATCAGCTTC
ATCTCAAGAAACACGTGAACATTTAGCCAAAAAGATTCTTCAATCTTACCACAAGTCTCT
CACCATCATGAACCTACTCCGCGCAACTTGACCAAGTTTCTCAGGGTGGAGGAAGCCCCAA
GAGCGATGATTCCGATCAAGAACCACCTTGTCATCAAGAGTTTCAAGAAGTCAATGCCAAG
GTGGAGTTCAAAAGTCAGAATTGCCCTGGAGCTGGTGTGATAGAACGCTGGACGATGG
ATTCAAGTTGGAGAAAGTACGGCCAGAAGGATATTCTCGGAGCCAAATTTCCAAGAGGATA
CTATAGATGCACGTATAGAAAGTCTCAAGGATGTGAAGCCACTAAACAAGTCCAAAGATC
TGATGAAAATCAGATGCTCCTTGAGATCAGTTACCGAGGAATACATTCTTGCTCTCAAGC
TGCAAAATGTCGGTACAACAATGCCGATACAAAACCTCGAACCGAACCCAGACCCAAGAACA
CGGAAATCTTGACATGGTAAAGGAAAGTGTAGACAACCTACAATCACCAAGCACATTTGCA
TCACAACCTTCACTATCCATTGTCTATCTACCCCAAATCTAGAGAATAACAATGCCTATAT
GCTTCAAATGCGAGATCAAAACATCGAATATTTTGGATCTACGAGCTTCTCTAGTGATCT
AGGAACTAGTATCAACTACAATTTTCCAGCATCTGGCTCGGCTTCTCACTCAGCATCAAA
CTCTCCGTCACCCGTCCTTTTGAATCCCCGTTTGAAAGCTATGATCCAAATCATCCATA
TGGAGGATTTGGTGGGTTCTATTCTTAGTTATCTACTTAAGGGAGGGACGGAACCTTTTA
CATGACCTCTTGATTAAAGAGAGAGTTTTTCATAATAGCTAATCAATTTCTATTCAAATA
TCCGAGTTTTTTTTCTAATCATGTTTATCAATTGTCTTATTACAGAAGGCTTATTTTCAG
GTCTATGTGAAATAAATGGATTGTACTCGTAGGTATGATCCTTGTATCTAAAAAAA
AAAAA

>G185 Amino Acid Sequence (domain in AA coordinates: 113-172)
MEKNHSSGEWEKMKNEINELMIEGRDYAHQFGSASSQETREHLAKKILQSYHKS LTIMNY
SGELDQVSQGGGSPKSDSDQEPLVIKSSKKSMPRWSSKVRIAPGAGVDR TLDDGFSWRK
YGQKDILGAKFPRGYRCTYRK SQCEATKQVQRSDENQMLLEISYRGIHSCSQANVGT
TMPIQNLBNQTEHGNLDMVKESVDNYNHQHLHNLHYPLSSTPNLENNNAYMLQMRD
QNIEYFGSTS FSSDLGTSINYNFPASGSASHSASNSPSTVPLESPFESYDPNHPYGGFGG
FYS*

>G1931 (5..592)
ATCAATGGAAGGGGTTGACAACACAAATCCTATGTTAACCTAGAGAAGGCGAAAACAA

CAATCCTTTTCTTCCTTAGATGACAAAACATTAATGATGATGGCTCCTTCGTAAATCTT
TTCGGGCGATGTAGGTCCATCTTCTTCTTGTACTCCAGCAGGTTATCATCTATCTGC
TCAGCTGGAGAACTTTTCGAGGAGGTGGAGGAGAGATGGGAGGATTAGTGAGTAATAATAG
CAATAATAGTGATCATAATAAGAATTGCAACAAAGGAAAAGGGAAGAGAACTTTGGCAAT
GCAGAGGATAGCTTTTCATACAAGGAGTGATGATGATGTTCTTGATGATGGTTATCGTTG
GCGAAAGTACGGTCAGAAATCTGTCAAGAACAATGCTCATCCAGGAGCTATTATAGATG
TACATACCACACATGCAACGTGAAGAAACAAGTGCAAGACTGGCAAAAGATCCAAACGT
TGTCGTAACAACTACGAAGGTTCATAATCATCCTTGAGAGAAGCTCATGGAGACTCT
TAGCCCTCTCCTTAGGCAACTTCAGTTCCTCTCAAGAGTTTCTGATCTGTAATTATTGAA
TGTTAATTAGTGGTGTAAATACATTAATTATGCTTTAATCTCTCCATTGACCCTCAATC
>G1931 Amino Acid Sequence (domain in AA coordinates: 114-170)
MEGVNDTNPMILTLEEGENNNPFSSLDKTLMMMAPSLIFSGDVGPPSSSSCTPAGYHLSAQ
LENFRGGGEMGGLVSNNSNNSDHNKNCNKGKGRITLAMQRIAFHTRSDDDVLDGGRWR
KYGQKSVKNNNAHPRSYRYCTYHTCNVKKQVQRLAKDPNVVVTTYEGVHNHPCEKLMETLS
PLLRLQLQFLSRVSDL*
>G2543 (1..2169)
ATGAGTTTCGTCGTCGGCGTGGCGGAAGTGGTAGTGAAGCGGCGGAGACGGTGGTGGT
AGTCATCATCACGACGGCTCTGAACTGATAGGAAGAAGAAACGTTACCATCGTCACACC
GCTCAACAGATTCAACGCCTTGAATCGAGTTTCAAGGAGTGTCCTCATCCAGATGAGAAA
CAGAGGAACCACTTAGCAGAGAATTGGGTTTGGCTCCAAGACAAATCAAGTTCTGGTTT
CAGAACAGAAGAACTCAGCTTAAGGCTCAACATGAGAGAGCAGATAATAGTGCATAAAG
GCAGAGAATGATAAAATTCGTTGCGAAAACATTGCTATTAGAGAAGCTCTCAAGCATGCT
ATATGTCCTAACTTGGAGGTCTCTCTGTAGTGAAGATCCTTACTTTGATGAACAAAAG
CTTCGGATTGAAAATGCACACCTTAGAGAAGAGCTTGAAGAATGTCTACCATGTCATCA
AAGTACATGGGAAGACCGATATCGCAACTCTCTACGCTACATCCAATGCACATCTCACCG
TTGGATTTGTCAATGACTAGTTTAACTGGTTGTGGACCTTTTGGTCAATGGTCTTCACTC
GATTTTGATCTTCTTCCAGGAAGTTCTATGGCTGTTGGTCTTAATAATAATCTGCAATCT
CAGCCTAACTTGGCTATATCAGACATGGATAAGCCTATTATGACCGGCATTGCTTTGACT
GCAATGGAAGAATTGCTCAGGCTTCTTCAGACAAATGAACCTCTATGGACAAGAACAGAT
GGCTGCAGAGACATTCTCAATCTTGGTAGCTATGAGAATGTTTCCCAAGATCAAGTAAC
CGAGGGAAGAACCAGAACTTTCGAGTCGAAGCATCAAGGTCTTCTGGTATTGTCTTCATG
AATGCTATGGCACTTGTGCGATGTTTCATGGATTGTGTCAAGTGGACAGAAGCTCTTCCC
TCTATCATTCGAGCTTCTAAAACACTTGCAGTGATTTCTTCAGGAATGGGAGGTACCCAT
GAGGGTGCATTGCATTTGTTGTATGAAGAAATGGAAGTGCTTTCGCCTTTAGTAGCAACA
CGCGAATTCTGCGAGCTACGCTATTGTCAACAGACTGAACAAGGAAGCTGGATAGTTGTA
AACGCTCTCATATGATCTTCTCAGTTTGTCTTCTCACTCTCAGTCTTATAGATTTCATCT
GGATGCTTGATTCAAGGATATGCCCAATGGATATTCCAAGGTACTTGGGTTGAACATATT
GAACTGAAGAAAAAGAACTGGTTTCATGAGCTATACAGAGAGATTATTACAGAGGGATT
GCTTTTGGGGCTGATCGTTGGGTTACCACTCTCCAGAGAATGTGTGAAAGATTGCTTCT
CTATCGGTACCAGCGTCTTCATCTCGTGATCTCGGTGGAGTGATTCTATACCGGAAGGG
AAGAGAAGCATGATGAGACTTGCTCAGAGGATGATCAGCAACTACTGTTTAAGTGTACGC
AGATCCAACAACACACGCTCAACCGTTGTTTCGGAAGTGAACGAAGTTGGAATCCGTGTG
ACTGCACATAAGAGCCCTGAACCAACCGGCACAGTCTTATGTGCAGCCACCCTTCTGG
CTTCCCAATTCTCTCAAAATGTCTTCAATTTCTCCTCAAAGACGAAAGAACCCGTCCTCAG
TGGGATGTTCTTTCAAACGGAACGCAAGTGAAGAAGTTGCTCACATCTCAAACGGATCA
CATCTTGAAACTGCATATCGGTTCTACGTGGATCCAATGCAACACATAGCAACAACATG
CTTATTCTGCAAGAAAGCTCAACAGACTCATCAGGAGCATTGTGGTCTACAGTCCAGTG
GATTTAGCAGCATTGAACATCGCAATGAGCGGTGAAGATCCTTCTTATATTCTCTCTTG
TCCTCAGGTTTCAATCTCACCAGATGGAAATGGCTCAAACCTCTGAACAAGGAGGAGCC
TCGACGAGCTCAGGACGGGCATCAGCTAGCGGTTCTGTTGATAACGGTTGGGTTTCAGATA
ATGTTAAGCAATTTACCGACGGCAAACTGAATATGGAGTTCGGTGGAAACGGTTAATAAC
CTGATAGGAACAACGTGATCAAAATTAACACCGCCTTGAGCGGTCTACAGCTTCAACT
ACAGCTTGA
>G2543 Amino Acid Sequence (domain in AA coordinates: 31-91)
MSFVVGVGGSGSGSGDGGGSHHDGSETDRKKRYHRHTAQQIQRLESSFKECPHPDEK
QRNQLSRELGLAPRQIKFWFQNRRTQLKAQHERADNSALKAEENDKIRCENIAIREALKHA
ICPNCGPPVSEDPYFDEQKLRIENAHLEELERMSTIASKYMGRPISQLSTLHPMHISP

LDLSMTSLTGCGPFGHGPSLDFDLPSSMAVGPNNNLQSQPNLAISDMDKPIMTGIALT
AMEELLRLQLTNEPLWTRTDGCRDILNLGSEYENVFPRSSNRGKNQNFVEASRSSGIVFM
NAMALVDMFMDVCVKWTELPFSIIAASKTLAVISSGMGGTHEGALHLLYEEMEVLSPLVAT
REFCELRYCQQTEQGSWIVVNVSYDLPQFVSHSQSYRFPSCGLIQDMPNGYSKVTWVEHI
ETEEKELVHELYREIIHRGIAFGADRWVTTLQRMCFERFASLSVPASSSRDLGGVILSPEG
KRSMMLRAQRMISNYCLSVSRNNTRSTVVSELNEVGIRVTAHKSPEPNGTVLCAATTFW
LPNSPQNVFNFLKDERTRPQWDVLSNGNAVQEVAHISNGSHPGNCISVLRGSNATHSNM
LILQESSTDSSGAFVVYSPVDLALNLIAMSGEDPSYIPLSSGFTISPDGNGSNSEQGGA
STSSGRASASGLITVGFQIMVSNLPTAKLNMESVETVNNLIGTTVHQIKTALSGETAST
TA*

>G264 (30..1430)

CTTGTACAGTTTCTGATTAGATTCAACAATGAACGGCGCATTAGGTAACCTCCTCCGCCT
CCGTTAGCGGCGGAGAGGAGCCGGAGGACCAGCGCTTTCTTGGTGAAAACCTACGAGA
TGGTCGACGATTCAACCGACAGATCGTATCGTGGAGCGCTAACAAACAGCTTCA
TCGTTTGAATCATGCCGAATTTTACGCCCTCCTTCTTCAACCTACTTCAAACACAATA
ACTTCTCTTCTTCTTCTCGTCAGCTCAATACCTATGGGTTTAGGAAGATTGATCCAGAGA
GGTGGGAGTTTTTGAATGATGATTTTATTAAGGATCAGAAGCATCTTCTCAAGAATATAC
ATAGAAGGAAACCTATACACAGCCACAGTCATCCACCTGCTTCGTCGACTGATCAAGAAA
GAGCAGTGTGCAAGAGCAAATGGACAAGCTTTCACGTGAGAAAGCTGCAATTGAAGCTA
AGCTTTTAAAGTTCAAACAACAGAAGGTTGTAGCAAAGCATCAGTTTGAAGAAATGACTG
AGCATGTTGATGATATGGAGAATAGGCAGAGAAGCTGCTGAATTTTTTGGAAACTGCCA
TTCGGAATCCTACTTTTTGTTAAGAATTTTGGTAAGAAAGTCGAGCAGTTGGATATTTTCA
CTTACAACAAAAGCGAAGGCTCCCTGAAGTTGAGCAATCAAAGCCACCTTCAGAAGATT
CTCATCTGGATAATAGTAGTGGTAGCTCGAGACGCGAGTCTGGAAACATTTTTTCATCAA
ATTTCTCTAATAAATTGCGACTAGAGCTTTCTCCAGCTGATTCAGATATGAACATGGTTT
CACACAGTATACAAAGTTCCAATGAAGAAGGTGCGAGTCCCAAAGGGATACTGTCAAGAG
GTGATCCAAATACTACACTAACAAAAGAGAAGGCCTACCATTTGCACCTGAAGCTCTAG
AGCTTGCGGATACCGGGACATGCCCGAGGAGATTACTGTTAAATGATAATAACAAGGGTGG
AGACCTTGACGACGAGGCTAACTTCTTCAGAGGAGACTGATGGTAGCTTTTCATGTCTAT
TAAATCTAACCTGGCTTCTGCTCCGTTACCGGACAAAACAGCTTCACAGATAGCTAAGA
CGACTCTTAAAGTCAAGAGTTAAACTTTAACTCAATAGAAACAAGTGCAAGTGAGAAAA
ATCGGGGTAGACAAGAGATTGCAGTTGGAGGTAGCCAAGCAAATGCAGCTCCTCCAGCAA
GAGTGAATGATGTATTCTGGGAACAGTTCCCTAACAGAAAGGCCAGGGTCTTCAGATAATG
AGGAGGCAAGTTCGACTTATAGAGGTAACCCATACGAAGAGCAAGAGGAGAAAAGAAACG
GGAGTATGATGTTACGTAATACAAAGAATATCGAGCAGCTGACCTTATAAACTATTTGGA
CGGTTACATCAACGAGAGTACGAACTGAGGTTTTGGTAAGAAGTATGGGTGAGTAAGTAA
TGAAACATTGGACTGAAAAGCGTAAGTAGCTTTGTTGTAAACACTTGCGTCTCTGTCTA
CACAAAGTAATTTGACTGTAATGTAAAGTTACAGGATTTAAATTGAATAAGCA

>G264 Amino Acid Sequence (domain in AA coordinates: 24-114)
MNGALGNSSASVSGGEGAGGPAPFLVKTYEMVDDSSDQIVSWSANNNSFIVWNHAEFSR
LLLPTYFKHNNFSSFIRQLNTYGFRKIDPERWEFLNDDFIKDQKHLKNIHRRKPIHSHS
HPPASSTDQERAVLQEQMDKLSREKAAIEAKLLKFKQKQVVAKHQFEEMTEHVDDMENRQ
KKLLNFLETAIRNPTFVNFGKKVEQLDISAYNKKRRLPEVEQSKPPSEDSHLDNSSGSS
RRESGNIFHQNFSNKLRLLESPADSDMNMVSHSIQSSNEEGASPKGILSGGDPNTTLTKR
EGLPFPAPEALELADTGTCPRLRLLDNDRVETLQORLTSEETDGSFCHLNLTLASAPL
PDKTASQIAKTTLSQELNFSIETSASEKNRGRQEIADVGSQANAAPPARVNDVFWEQF
LTERPGSSDNEEASSTYRGNPYEEQEEKRNGSMMLRNTKNIEQLTL*

>G32 (101..736)

AACACACATTCCCTCTCTTCTTCAACTAGAAAAAGATAGATATATCGGACATTTATTG
ATCTGTGTATGCATAAAGGTATAGTATCATTTATTAGAAAGATGAACACAACATCATCAA
GAGCAAGAAGAAGCAAGACGATCAGGTTGGTACAAGGTTTCTTGGGGTGAGAAGAAGGCC
TTGGGGAAGATACGCAGCTGAGATTAGAGACCAACTACGAAGGAGCGTCACTGGCTTGG
CACTTTCGATACGGCGGAAGAAGCTGCCTTGGCTACGATAGAGCTGCTCGGTCCATGCG
TGCCACACGTGCCGACGAACCACTTTGTTTACTCAGACATGCCTCCTTCTCATCCGTCAC
CTCCATTGTTTCTCCTGACGATCCTCCTCCTCCTCCACCTCCTCCTGCTCCTCCTAGCAA
TGATCCTGTGATTACATGATGATGTTTAACCAATACTCATCCACTGACTCGCCAATGCT
TCAGCCTCATTGTGATCAAGTGGACAGTTACATGTTTGGTGGCTCTCAATCTTCGAATTC

TTATTGCTATTCTAATGACAGTAGTAATGAGCTGCCTCCTCTCCCAGCGACTTGTGAA
TTCGTGTTATAGCCAACCACAGTGGACCTGGACCGGTGACGACTACTCGTCTGAGTACGT
ACATAGTCCAATGTTTCAGCAGAATGCCTCCGGTTTCTGACTCTTTCCCTCAAGGTTTCAA
CTACTTTGGCTCCTAATTCTTTCTCATCGTCCATATTTAATACCTTCCTCATTGTGACCT
TTTCTCTTCTTCTTTTTTGGGTTTATCTATGTTTCGCCGCTCTTGATCTCTGCCATATG
TGATCAAAGTGACTGTTGTCAATTAGTTTTTCAATAACAAGTTATCATTGTATCTTGAA
AAAAAAAAAAAA

>G32 Amino Acid Sequence (domain in aa coordinates: 17-84)
MNTTSSKSKKKQDDQVGTRFLGVRRRPWGRYAAEIRDPTTKERHWLGTFDTABEEAALAYD
RAARSMRGTRARTNFVYSMDPPSSSVTSIVSPDDPPPPPPPPAPPSPNDPVDYMMMFNQYS
STDSPMLQPHCDQVDSYMFSGQSNNVSYCSNDSSNELPPLPSDLNSNCYSQPQWTWTGD
DYSSYVHSPMFSMPFVSPVSDSFPQGFNYFGS*

>G436 (1..2157)
ATGGATTTTACTCGCGATGACAACCTCAAGTGATGAACGGGAAAATGATGTAGACGCCAAC
ACCAACAACCGTCACGAGAAGAAGGGTTACCATCGCCACACTAATGAACAAATTCATAGG
CTTGAAACGTATTTCAGGAATGTCTCATCCAGACGAATTTAGCGACGCTCTGTTGGGT
GAAGAAGCTGAATCTGAAACAAAACAAATCAAATTTTGGTTTCAAACAAAAGAACTCAA
GCTAAGAGTACCAATGAAAAGCAGACAATGCAGCGCTTAGGGCAGAAAATATTAAGATT
AGACGTGAGAAGCAATCAATGGAAGATGCACTGAATAATGTGGTTTGGCCCTCCATGTGGT
GGTCGTGGTCTCTGGGAGAGAAGACCAACTTCGACATCTCCAAAACTCCGTGCACAAAAC
GCTTATCTCAAAGATGAGTATGAAAGAGTCTCAAACCTACCTAAAACAGTACGGAGGTCAC
TCAATGCATAACGTGAGGCCACACCTTATCTCCATGTGTCCATCAAACCATGCATCAACG
TCCAAGAACCGTCCAGCATTGTACGGAACCTCTTCTAACCGTCTCCCCGAGCCTTCAAGC
ATATTTAGAGGACCATACTCGTGGAAACATGAACACACCCGACCCGCTCAGCCGCGA
AAGCCGCTGGAAATGCAGAATTTCCAACCACTATCTCAACTGGAGAAAATTCGAATTTGG
GAAGCAGCGGAAAAGCGGGTGTAGAGGTTTGTAGCCTCATTCAAATGGATGATACAATG
TGGAAAAAGTCGTCTATTGATGATAGGCTCGTCATTGATCCAGGGCTCTATGAGAAATAT
TTTACTAAGACTAACACAAATGGTCTGCTCTGAGTCTTCTAAAGATGTCTGTGGTGGTTCAA
ATGGATGCTGGAACTTGATCGACATCTTCTTAACTGCGGAGAAAATGGGCGAGGCTTTTT
CCAACAATTGTGAACGAAGCTAAAACGATTACGTCCTTGGATTCCGTTGACCATCGAGGA
AAAACTTTCTCAAGAGTGATTATGAGCAACTGCACATACTGTCAACATTTGGTGCCACCG
AGGGAATTTATGATCCTAAGGACTTGCCAAACAAATGGAAGCAATGTCTGGATGCTGT
GATGTGTCTGTGCTATCCCAAACATTGAGTTTGTATCTTTCTGTTTCCCATTTGCACAAA
GTCCTCCTCAGGTGTGCTCATTCAAGCCTTGCCCCACGGCTTCTCTAAGGTGACGTGGATA
GAGCATGTGGTAGTGAATGATAATAGAGTGCGGCCACATAAGCTTTACAGAGACCTCTTA
TACGGCGGGCTTTGGCTACGGAGCTCGACGTGGACCGTTACTCTTGAGAGGACGTGTGAG
AGGCTGATTTTCTCCACCTCCGTCCCTGCTTGCCCAACATGACAATCCCAGGAGTTGTG
CAACAATACGAGGCAGAAAATAGCGTAATGCATTTGGGAGAAAAGATGTTGAGGAACATT
GCATGGATGATGAAAAATGGTTAAACAAACTCGACTTCTCGCCACAGTCTGAACTAACAC
AGCGGAATTAGGATTGGGGTGGGATAACAACTAGGGCGGTCACACCGCCGCTCTCATTT
GCTGTGCTGGTTTCTATCTTTATCCCTCCCTCTCCCTCCTGTCCAAGTGTACGATTTCTCT
AAGAATCTGGAGGTTCTGTCACCAGTGGGACGTTCTGTGCCATGGGAATCCAGCGACTGAG
GCTGCTCGTTTCTGTCACCGGATCAAACCCAAGGAACACTGTGTCTTTTCTCGAGCCTTCA
ATTAGGGATATTAATACTAAGCTAATGATACTCCAAGATAGCTTCAAAGATGCATTGGGA
GGAATGGTGGCCTACGCTCCAATGGATCTAAACACCGCCTGCGCTGCCATTTTCAGGCGAT
ATCGATCCTACCACCATTTCCAATCTCTCCCTTCCGGTTTTATGATCTCCCGTGACGGCCGT
CCTTCCGAGGGCAGCGGAGGGTGGCAGCTATACACTCTCACCGTGCTTTCCAGATC
CTTGTCCTCGGTCCSAGTTACTCTCTGTATACCAACCTGGAGGTTTCTGCCACACAGTC
AATACTTTGATTAGCTCCACCGTTCAAAGGATCAAAGCCATGCTCAAGTGCGAATGA

>G436 Amino Acid Sequence (domain in AA coordinates: 22-85)
MDFTRDDNSSDERENDVDANTNNRHEKKGYHRHTNEQIHRLETYFKCEPHPDEFQRRLLG
EELNLKPKQIKFWFQNKRTQAKSHNEKADNAALRAENIKIRRENESMEDALNNVCPPCG
GRGPGREDQLRHLQLKLRQANAYLKDEYERVSNYLKQYGGHSMHNVEATPYLHGSPSNHAST
SKNRPALYGTSSNRLPEPSSIFRGPYTRGNMNTAPPQPRKPLEMQNFQPLSQLEKIAML
EAAEKAVSEVLSLIQMDTTFWKKSSIDRLRVIDPGLYEKYFTKTNINGRPESSKDQVVVQ
MDAGNLDIDFLTAEKWARLFTPTVNEAKTIHVLDSDVHRGKFTSRVIEQLHILSLPLVP
REFMILRTCOOIEDNVMMIADVSHCLPNIEFDLSFPICTRKPSGVLIOALPHGFSKVTWI

EHVVVNDNRVRPHKLYRDLLEYGGFGYGARRWTVTLERTCERLIFSTSVPALPNNDNPGVV
QTIRGRNSVMHLGERMLRNFAMMMKMNKLDLDFSPQSETNNSGIRIGVRINNEAGQPPGLI
VCAGSSLSLPLPPVQVYDFLKNLEVRHQWDLCHGNPATEAARFVTGSNPRNTVSFLBPS
IRDINTKLMILQDSFKDALGGMVAYAPMDLNTACAAISGDIDPTTIPILPSGFMISRDR
PSEGBAEGGSYTLTVAFFQILVSGPSYSPDTNLEVSATTVTNTLISSTVQRIKAMLKCE*

>G556 (50..1144)

CTTTTTTGAAGCCCTTTTGACACAAAAGACCAGAACAAAGTTGAAGAAATATGAATACAAC
CTCGACACATTTTGTTCACCGAGAGGTTTGAAGTTTACGAGCCTCTCAACCAAATCGG
TATGTGGGAAGAAAGTTTCAAGAACAAATGGAGACATGTATACGCCTGGCTCTATCATAAT
CCCGACTAACGAAAAACCAGACAGCTTGTCTAGAGGATACCTTCTCATGGGACAGAAGGAAC
TCCTCACAAAGTTTGACCAAGAGGCTTCCACATCTAGACATCCTGATAAGATACAGAGAAG
GCTAGCACAGAACTCGAGAGGCAGCTAGGAAAAGTCGTTTGCAGCAAGAAAGCTTATGTTCA
GCAGCTAGAGACTAGCCGGTTAAAGCTAAATTCATTTAGAGCAAGAACTCGATCGTGCTAG
ACAACAGGGTTTTCTATGTGGGGAACGGAGTAGATACCAATGCTCTTAGTTTTCTCAGATAA
CATGAGCTCAGGGATTGTTGCATTTGAGATGGAATATGGACATTGGGTGGAAGAACAGAA
CAGGCAAATATGTGAACTAAGAACGGTTTTACATGGACAAGTTAGTGATATAGAGCTTCG
TTCTCTAGTCGAGAATGCCATGAAACATTACTTTCAACTCTTCCGAATGAAGTCAGCCGC
TGCAAAAATCGATGTTTTCTATGTCTATGTCGGAATGTGGAAAACCTTCAGCAGAGCGGTT
TTTCTTGTGGATAGGCGGATTTAGACCCCTCAGAGCTTCTCAAGGTTCTGTTACCGCATTT
TGATCCTTTGACGGATCAACAACCTTTTGGATGTATGTAATCTGAGGCAATCATGTCAACA
ATCAGAAGATGCGTTATCCCAAGGTATGGAGAAAACCTGCAACATACATTAGCAGAGAGTGT
AGCAGCCGGGAAACCTTGGTGAAGGAAGTTATATTCTCAAATGACTTGTGCTATGGAGAG
ATTGGAGGCTTTGGTCAGCTTTGTAAATCAAGCTGATCATCTGAGACATGAGACATTGCA
ACAGATGCATCGGATCTTAACCACGCGACAAGCGGCTAGAGGTTTGTAGCATTAGGGGA
GTATTTCAAAGGCTTCGAGCTTTGAGTTCGAGTTGGGCGGCTAGGCAACGTGAACCAAC
GTAATTAAGGTGTTTAGATGTCAAGAAAGGTTTGAACCTTAACAATCAAGAATGGAGTT
TGCTGGTGAGTGGAATTTTGGGTCAAGAACAGAGCAATAACACAAGCTGCTGTGTGATG
ATGAATCTTGTCTTGGCGCTAAAGGAAATGTTTGAAGAAAGTTGTACATATGATCAGCAA
CGTAAAGTTTATAGCTTTTGAAGAACCAACTTTTCGATGGTTGTTCTTTTTTTTTTGTAT
GTAATATTATAGATAAGCTTGTGGTATATATGATTTTAATGTGACATTACGAACCTTGATT
TATAACCATGGTAAAT

>G556 Amino Acid Sequence (domain in AA coordinates: 83-143)

MNTTSTHFVPPRRFEVYEPLNQIGMWEESEFKNNGDMYTPGSIIIPTEKPDLSLSEDTSHG
TEGTPHKFDQEAESTSRHPDKIQRRLAQNREARKSRLRKKAYVQQLBTSRLKLIHLEQEL
DRARQQGFYVNGVDTNALSFSDNMSSGIVAFEMEYGHWVEEQNRQICELRTVLHGQVSD
IELRSLVENAMKHYFQLFRMKSAAAKIDVFYVMSGMWKTSAERFFLWIGGFRPSELLKVL
LPHFDPLTDQQLLDVCNLRQSCQSEDALSQGMKQLQHTLAESVAAGKLGEYSYIPQMT
AMERLEALVSVFNQADHLRHETLQQMHRILTTRQAAGLLALGEYFQRLRALSSSWAARQ
REPT*

>G1420 (39..1238)

AAAGTATCATCTCATAGATTCATCTTTTCTCTATTACATGGAGAAGAAAAAGAAGAGG
ATCATCATCATCAACAACAACAACAACAAGGAGATCAAGAACACAGAGACAAAGA
TCGAGCAAGAACAGAACAAGAACAAAAACAAGAAATCTCTCAAGCATCATCATCAA
ACATGGCGAATCTAGTTACGTCATCAGATCATCATCCGTTGGAGCTAGCTGGAAATCTCT
CAAGCATCTTCGATACTTCATCTTTACCTTTTCTTATTCTTATTTTGAAGATCACTCTT
CTAATAATCCTAATTCTTTCTTAGACTTGCTCCGACAAGATCATCAGTTTGCTTCTTCT
CTAATTCCTCTTCTTTTTCATTTCGATGCCTTTCTCTCTCCCAATAACAACAACAACCT
CTTTTTTTACGGATTTGCCCTTACCTCAAGCTGAGTCATCAGAAGTCGTGAACACAACAC
CGACTTCTCCAACTCAACCTCAGTCTCATCTTCTCCAACGAAGCTGCAAATGATAACA
ACAGTGGTAAAGAAGTTACTGTTAAAGATCAAGAAGAAGGAGATCAACAACAAGAGCAAA
AGGGTACTAAGCCACAGTTGAAGGCAAGAAGAAGAAATCAAAAGAAAGCTAGAGAAGCTA
GGTTTGGCTTTCTGACGAAGAGCGATATTGATAATCTTGACGACGGTTATAGGTGGAGAA
AATACGGCCAAAGCTGTCAAAAACAGTCCTTATCCAGAAGCTATTACCGTTGCACCA
CAGTGGGTTGCGGAGTGAAGAAGAGAGTGGAGAGATCCTCCGATGATCCTTCGATCGTCA
TGACAACCTACGAAGGTCAGCATACCCATCCTTTCCCATGACGCCACGTGGACACATCG
GAATGCTCAGGTCACCAATCCTAGACCACGGTGCAACCACCGGTCATCATCATCATTCT
CCATCCCTCAGCCACGTTACTTGCTGACTCAACATCACCAGCCCTACACATGTACAACA

ACAACCTCTCTAAGTATGATCAATAGAAAGATCATCCGATGGCACTTTTCGTAAATCCAGGTC
CATCATCATCATTTCCCCGGCTTTGGTTATGATATGTCTCAAGCTTCTACTTCAACTTCTT
CTTCCATTAGAGATCATGGATTGCTTCAAGATATTCTTCCTTCGCAGATCAGATCCGATA
CTATTAACACTCAAACCAATGAAGAGAATAAGAAATGAAGAAGTTTTTTTTTCCCGGGGCA
ATTGTTTTTTTTCTTTAGGCCGGATCCGGTAGGTAGGTTTCATGAGC
>G1420 Amino Acid Sequence (domain in AA coordinates: 221-280)
MEKKKEEDHHHQQQQQQQKEIKNTETKIEQEQEQEQEQEISQASSSSNMNANLVTSSDHP
LELAGNLSIFDTSLLPFPYSYFEDHSSNNPNSFLDLLRQDHQFASSSNSSSFSFDAFPL
PNNNNNTSFFDLPPLQAESSEVVNTTPTSPNSTSVSSSSNEAANDNNSGKEVTVKDQEE
GDQQEQEQGTPQLKAKKKNQKKAREARFAFLTKSDIDNLDGYSRWRKYGQKAVKNSPYP
RSYYRCTTVGCGVKRVERSSDDPSIVMTTYEQHHPFMTPRGHIGMLTSPILDHGAT
TASSSSFSIPQPRYLLTQHHQPYNNMNNLSMINRRSSDGTFFVNPFPSSSPGFGYDMS
QASTSTSSSIRDHGLLDILPSQIRSDTINTQTNEENKK*
>G1412 (115..1008)
CCCACGCGTCCGCCACGCGTCCGAAACAAAAACATATAATTTGGGTTTTTAGAGTTCTGA
AACTTGAAATCTTTTTTTTTTTGGTTGCTGAGGAATCGAAGTAGAAGAGTATAAATGGGT
GTTAGAGAGAAAGATCCGTTAGCCAGTTGAGTTTGCCACCAGGTTTTAGATTTTATCCG
ACAGATGAAGAGCTTCTTGTTCAATCTATGTCGGAAGTTGCAGGCTATCATTTCTCT
CTCCAGGTCATCGGAGACATCGATCTCTACAAGTTTCGATCCTTGGGATTTGCCAAGTAAG
GCTTTGTTTGGAGAGAAGGAATGGTATTTCTTTAGCCCAAGAGATCGGAAATATCCGAAC
GGGTCAAGACCCAATAGAGTAGCCGGTTCGGGTATTGGAAAGCAACGGGTACTGACAAA
ATTATCACGGCGGATGGTCGTGTCGGGATTAAAAAAGCTCTGGTCTTTTACGCCGGA
AAAGCTCCCAAGGCACTAAAACCAACTGGATTATGCACGAGTATCGCTTAATAGAACAT
TCTCGTAGCCATGGAAGCTCCAAGTTGGATGATTGGGTGTTGTGTCGAATTTACAAGAAA
ACATCTGGATCTCAGAGACAAGCTGTTACTCTGTTCAAGCTTGTCTGAAGAGCATAGC
ACGAATGGGTCTGTCATCGTCTTCTTCATCAGCTTGACGACGTTCTTGATTCTGTTCCCG
GAGATAAAAGACCAGTCTTTAATCTTCTCGGATGAATTCGCTCAGGACGATTCTTAAC
GGGAACTTTGATTGGGCTAGCTTGGCAGGTCTTAATCCAATTCAGAGCTAGCTCCGACC
AATGGATTACCGAGTTACGGTGGTTACGATCGGTTTCGAGCGCGGAAGGTGAGGCGGAG
AGTGGGCATGTGAATCGGCAGCAGAACTCGAGCGGGTTGACTCAGAGTTTCGGGTACAGC
TCGAGTGGGTTTTGGTGTTCGGGTCAAACATTGAGTTTAGGCAATGAGAGAGATGTGAA
GTTACTGATGGGTGAAAAAAGTAAAAAAGAACTTGGAGATAGTAGAGTGCAATTGATG
TAAATAATAGGGATTTATATGGGGCTTTTACCGATTGCGGTGAGGCTTAGGATTCCCAAA
GGAAAAGGCTCGACTGGGACTAGTTTGATCCAACCTGACGGCCCCCAATGTGTAAATG
TTCTCAACGGAGAGAAAAATAAATGGTTACCAATATTTTTCCAAAAAAGAAAAA
>G1412 Amino Acid Sequence (domain in AA coordinates: 17-159)
MGVREKDLAQLSLPPGFRFYPTDEELLVQYLCRKVAGYHFSLQVIGDIDLYKFDPWDL
SKALFGEKEWYFFSPDRKYPNGSRPNRVAGSGYWKATGTDKIITADGRRVGIIKKALV
AGKAPKGTKTNIWIMHEYRLIEHSRSHGSSKLDDWVLCRIYKKTSGSQRQAVTPVQACRE
HSTNGSSSSSSQLDDVLSDFPEIKDQSFNLPRMNSLRITLNGNFDWASLAGLNPIPELA
PTNGLPSYGGYDAFRAEAGEAESHVNRQQNSSGLTQSFYSSSGFVSGQTFEFRQ*
>G738 (1..885)
ATGGACCATCATCAGTATCATCATCATGATCAATACCAACATCAGATGATGACTAGTACT
AACAATAATTCCTATAACACCATCGTCACAACACAACCACCACCAACAACAACAATG
GATTCACAACACGACAACAACTATGATAATGGATGACGAGAAGAAGTTGATGACGACAATG
AGCACTAGGCCCGCAAGAACCAAGAACTGTCCAAGATGCAACTCAAGCAACACCAAGTTT
TGTATTACAACAACATACAGCTTAGCACAGCCTAGGTACTTGTGTAAGTCTTGTGCGGAGA
TATTGGACTGAAGGFGGCTCTCTCCGTAACGTCCCCGTAGGCGGAGGTTCTAGAAAGAAC
AAGAAGCTTCCATTTCTTAATTCCTCTACTTCTTCTTCCACCAAGAACCTCCCGGATCTC
AACCCTCTTTCTGCTCTTACATCATCAGCTTCTCATCAAAACCTTAGCAAGACGCATCAA
AACAATAATGACCTCAGCCTATCCTTCTCTTCCCTATGCAAGACAAGCGAGCTCAAGGG
CATTACGGTCATTTTCAGTGAGCAAGTTGTGACAGGAGGGCAGAACTGTCTTTTCAAGCT
CCTATGGGAATGATTAGTTTCGTCAAGAGTATGATCATGAGCACCACCAAGAAATCTT
GGGTTTTTCATTAGACAGGAACGAGGAAGAGATTGGTAATCATGATAACTTCGTTGTTAAT
GAGGAAGGAAGTAAGATGATGATCCTTATGGAGATCATGAAGACCGTCAACAACATCAC
CATGTGAGACACGATGATGGTAATAAGAAGAGAGAAGGTGGTTCAAGCAATGAGCTATGG
AGCGGAATCATCCTAGGTGGTGATAGTGGTGACCAACATGGTGA

>G738 Amino Acid Sequence (domain in aa coordinates: 351-393)
MDHHQYHHHDQYQHQMSTNNNSYNTIVTTQPPPTTTTMDSTTATTMIMDDEKKLMTTM
STRPQEPNRCPRCNSSNTKFCYNNYSLAQPRYLCKSCRRYWTEGGSLRNVFVGGGSRKN
KKLPFPNSSTSSSTKNLPDLNPPFVFTSSASSSNPSKTHQNNNDLSLSFSSPMQDKRAQG
HYGHFSEQVVTGGQNCFLQAPMGMIQFRQYDHEHPKKNLGFSLDRNEEEIGNHNDNFVVN
EEGSKMMPYPYGDHEDRQHHHVRHDDGNKKREGGSSNELWSGIILGGDSGGPTW*

>G2426 (1..1038)

ATGGGCAGATCGCCATGTTGTGATAAGGCCGGGTGAAGAAAGGCCTTGACTCCAGAA
GAGGATCAGAACTTTTGGCTTATATTGAAGAACATGGCCATGGAAGCTGGCGTTCTTTG
CCTGAGAAAGCCGCTCTCCAAAGGTGTGGAAGAGTTGCAGACTCAGATGGACTAACTAC
CTAAGACCTGACATCAAGAGAGGCAAATTAAGTGTACAAGAAGACAAACCATCATTCAA
CTCCACGCTCTCTCGGAAACAGGTGGTCAGCGATTGCAACTCATTTACCAAAGAGGACA
GACAACGAGATCAAGAACTACTGGAACACACACTTGAAGAAACGCTGTATCAAAATGGGG
ATAGATCCAGTGACTCACAAGCACAAAACGAGACTCTTTCGTCTTCCACAGGACAATCA
AAGAACGCAGCCACGCTTAGTCATATGGCTCAATGGGAGAGTGCAAGACTCGACGCTGAA
GCAAGGCTAGCTAGAGAATCAAAGCTTCTCCATTTACAGCATTACCAAAACAATAACAAC
CTTAACAAATCAGCAGCTCCTCAACAACATTGCTTCACTCAAAAAACATCAACAACTGG
ACTAAACCAAACCAAGGAAACGGAGACCAACAGCTTGAATCTCCGACATCGACGGTGACA
TTCTCTGAGAATCTTCTGATGCCTTTAGGAATCCCTACGGATAGCAGCAGAAATAGAAAC
AATAACAACAATGAGTCTCGCGCATGATTGAATTGGCCGTATCTTCGTCAACCTCCTCC
GATGTGAGTCTGGTCAAAGAACATGAACAGACTGGATTAGGCAGATCAACTGTGGTAGT
GGAGGAATAGGAGAAGGATTACAGAGTCTATTGATCGGTGATTCGGTTCGGCCGGGGTTA
CCCAACGGGAAAAACGAAGCGACGGCGGGCGTGGGAATGAGAGTGAGTATAACTACTAT
GAGGATAACAAGAATTACTGGAATAGCATTCTCAACTTGGTTGATTCTTCACCGTCCGAT
TCCGCGACGATGTTCTGA

>G2426 Amino Acid Sequence (conserved domain in AA coordinates:14-114)

MGRSPCCDKAGLKKGPWTPPEEDQKLLAYIEEHGHGSWSRLPEKAGLQRCGKSCRLRWNY
LRPDIKRGKFTVQEEQTIQLHALLGNRWSAIATHLPKRTDNEIKNYWNTHLKKRLIKMG
IDPVTHKKHNETLSSSTGQSKNAATLSHMAQWESARLDAEARLARESKLLHLQHYQNNNN
LNKSAAPQQHCFQTQKTSNWTNTPNQNGDQQLSEPTSTVTFSENLLMPLGIPTDSSRNRN
NNNNESSAMIELAVSSSTSSDVSILVKEHEHDWIRQINCGSGGIGEGFTSLLIGDSVGRGL
PTGKNEATAGVGNESYNYEDNKNYWNLSILNLVDSSPSDSATMF*

>G1524 (1..825)

ATGGGGAGAACTAAGGAGCAGGCAACATTAAGTTCGGTATCCACCTGTCTAGGAATCCT
GCTAAATTCATGATATAAAACAAAGCACTCCAGGAAAAAGGATATGGTAAGGCTCTGAAA
AGAAAACCTTGGACGGGTGTGACATGCCCTGTCTGTCTTGAGGTTCTCACAACTCGGTC
GTCCTCTCTTTGTTTATCTTACCACAAAGGATGCCGTCCGTACATGTGTGCCACGGGAAAC
CGTTTCTCAAATTTGTCTAGAGCAGTACAAAAGGCATATGCCAAGGATGAGAAAAGTGAC
AAACCGCCAGAGCTATTGTGCCCCGCTTTGTAGGGGTGAGGTGAAAGGCTGGACCGTTGTG
GAAAAGGAACGTAAGTATCTGAATCTAAGAAAAGGTGATGCATGAACGACGAGTGTGTTG
TTTTATGGAAGCTATAGACAGCTCAAGAAGCATGTTAAGGAGAACCATCCGAGAGCCAAG
CCAAGAGCCATAGACCTGTGCTGGAGGCGAAATGGAAGAAGCTTGAGGTTGAGAGGGAG
AGGAGTGATGTAATCAGCACAGTCATGTCTCAACACCTGGGGCTATGGTATTTGGAGAC
TATGTGATTGAGCCATACATGGTTATGATCATCAAGATGACAGTGACGATTACAGTGAT
TCGTCGGATGACGAAATGGAAGGTGGGGTATTTCGAGCTTGAGAGCATTCGACCTGGGCCGT
CTTCAACCGCGCTTCGGCTGCCATCTCAAGCCGGGGAATTCGCGGTATGATCATAAGGAAC
CGGTGGGCTCGAAGCAGAGGTGCGAGCAGAAGCGCACAACATAA

>G1524 Amino Acid Sequence (conserved domain in AA coordinates:49-110)

MGRTEKEQATLTRYPPCPNPAKFNDINKALQEKYKALKRKPWTGVTCPVCLEVPHNSV
VLLCSSYHKGCRPYMCATGNRFSNACLEQYKKAYAKDEKSDKPELLCPLCRGQVKWTVV
EKERKYLNSKKRSCMNDECLFYGSYRQLKKHVKENHPRAPRAIDPVLEAKWKLEVERE
RSDVISTVMSSTPGAMVFGDYVIEPYNGYDHQDDSDSDSDDDEMEGGVFELGAFDLGR
LQPRSAAISSRGIRGMIIRNRWARSRGASRRRQT*

>G1243 (1..3174)

ATGGCGAGAAATTCGAATCCGATGAGGCTTTCTCGTCAGAGGAGGAAGAAGAGCGGGTT
AAGGATAATGAAGAAGAAGATGAGGAGGAGCTCGAGGCTGTTGCTCGTCTTCTGGCTCC
GACGATGACGAAGTAGCCGCCCGACGAATCACCAGTCTCCGACGAGAGGCTGCTCCC

GTAGAAGATGATTACGAGGACGAAGAAGATGAGGAAAAAGCTGAAATCAGCAAACGTGAG
AAAGCCAGACTTAAAGAGATGCAGAAGTTGAAGAAGCAGAAGATTCAAGAGATGCTGGAG
TCGCAGAATGCTTCCATTGACGCGGATATGAACAATAAGGGAAAAGGGAGACTGAAGTAT
CTTCTGCAGCAAACGTAGTTATTTGCCCACTTTGCTAAAAGTGATGGATCTTCTTCTCAG
AAGAAGGCAAAAGGAAGGACGTCATGCTTCCAAAATAACTGAAGAGGAGGAAGACGAA
GAGTATCTAAAGGAAGAAGAGGATGGCTTAACCTGGATCTGGAAACACACGGTTACTCACA
CAGCCCTCTTGATTCAAGGGAAGATGAGAGATTACCAATTAGCTGGTTTGAAGTGGCTC
ATTCGTCTTTATGAGAATGGCATAAATGGAATCTTGCTGATGAAATGGGTCTGGGGAAG
ACGCTTCAAACGATTTCTTTGTGGCATATCTTCATGAATACAGGGGAATCAATGGTCCC
CATATGGTGGTTGCTCCAAAATCAACACTTGGTAATTGGATGAACGAAATTCGCCGGTTT
TGTCTGTCTTACGTGCTGTGAAGTTTCCTTGGTAATCCTGAGGAGAGGAGACATATTCGA
GAAGACCTGCTAGTTGCTGGGAAATTTGATATTTGTGTCAAGCTTTGAGATGGCCATC
AAAGAGAAGACAGCACTTCGTCGGTTTAGCTGGCGTTATATTATCATTGATGAAGCGCAT
CGAATCAAGAACGAGAATTCACCTCCTTTCTAAAACCATGAGACTTTTGTAGCACAATTAT
CGGCTTCTTATCACGGGGACCCCCCTTCAGAATAATCTCCATGAAGTGTGGGCTCTTCTA
AATTTTCTTCTGCTGAGATTTTGTAGTTGAGCAGAGACTTTTGTATGAATGGTTTCAAATT
TCTGGTGAGAATGACCAGCAAGAAGTTGTGCAACAACGCAAGGTTCTTCGACCATTT
CTTCTTCGAAGACTAAAGTCAGATGTTGAGAAAGGTTTGCCACCGAAGAAGGAGACCATA
CTTAAAGTTGGTATGTCTCAGATGCAAAAGCAATACTACAAGGCTTTACTGCAGAAGGAT
CTTGAAGCGGTTAATGCTGGTGGGAGAACGCAACGCTGCTAAACATGCAATGCAACTG
CGTAAATGCTGCAATCACCCCTATCTCTTCCAGGGTGCAAGCTGGTCCCCCATATACC
ACAGGAGATCACCTTATAACAAAATGCTGGTAAGATGGTTCTCTTGGATAAATGCTTCTCCT
AAGTTGAAAGAAGCTGATTCAAGGGTGCTGATATTTTCTCAGATGACAAGACTTTTGGAT
ATTCTTGAGGACTATTTAATGTATCGTGGTTACTTGTATTGCCGTATTGATGGAAACACT
GGTGGTGACGAACGAGATGCCCTCCATAGAAGCCTACAACAAGCCAGGAAGTGAGAAATTT
GTTTTCTTGTATCTACTAGAGCTGGAGGGCTTGGTATCAATCTTGCCTACTGCAGATGTT
GTGATCCTTTACGATAGTGATTGGAACCCACAAGTCGACTTGCAAGCTCAGGATCGTGCC
CATAGGATTGGTCAAAAAAAGAAGTTCAAGTGTTCGATTCTGCCTGAGTCTGCTATT
GAGGAGAAAGTGATTGAAAGAGCTTACAAGAAGTTAGCACTTGATGCTCTGGTTATTCAA
AAGGGAGATTGGCAGAACAGAAAAGTAAGTCTGTCAATAAGGATGAGTTGCTTCAAATG
GTAAGATATGGTGCTGAGATGGTGTTCAGTTCTAAAGATAGCACAATCACAGACGAGGAT
ATTGATAGAATCATTGCCAAAGGAGAAGAGGCAACAGCTGAACTTGATGCTAAGATGAAG
AAATTCACAGAAGATGCTATACAGTTTAAATGGATGACAGTGCTGACTTCTATGATTTT
GATGATGACAATAAGGATGAAAACAAGCTCGATTTTAAAAGATTGTAAGCGACAATTGG
AATGATCCCCCAAGCGGGAGAGAAAGCGCAACTACTCTGAATCTGAGTACTTTAAGCAA
ACATTGCGGCAAGGTGCTCCAGCTAAACCTAAAGAGCCTAGAATTCCGCGCATGCCCCAG
TTGCACGATTTCCAGTTCTTTAACATTGAGAGATTGACCGAGTTGTATGAAAAGGAAGTA
CTTATCTCATGCAACATCAGAAAAATCAGTTGAAAGACACAATTGATGTTGAAGAA
CCAGAAGTTGGGATCCCTTAACCTACTGAAGAAGTAGAAGAAAAGGAGGGATTATTGGAG
GAGGGTTTCTCAACATGGAGCAGAAGAGATTTTAATACTTTCTCAGGGCTTGTGAGAAG
TATGGCCGCAACGACATAAAAAGCATTGCCTCTGAGATGGAAGGGAAAACAGAGGAAGAA
GTTGAAAGATATGCCAAAGTATTTAAAGAGCGGTACAAGGAGCTGAACGACTATGATAGA
ATCATTAAAGAACATTGAGAGGGGAGAGGCAAGGATCTTAGGAAAGACGAAATCATGAAG
GCCATAGGGAAGAACTGGATCGCTACAGAAACCTTGGCTGGAAGTGAAGATTCAATAT
GGTCAGAACAAAGGCAAGCTGTACAATGAAGAGTGTGACCGTTTCATGATCTGCATGATT
CACAAAATTGGTTATGGGAATTGGGATGAGCTAAAGGCAGCATTTAGGACATCGTCTGTG
TTCAGGTTTGACTGGTTTGTGAAATCCCGCACGAGTCAGGAACCTTGCAAGAAGATGCGAC
ACTCTGATTGCTGATCGAGAAAGAGAACCAGGAGTTTGATGAAAGAGAGAGGCAAGCC
CGCAAAGAGAAGAAGCTCGCGAAGAGTGCAACACCATCAAAGCGACCTTTAGGAAGACAA
GCAAGTGAGAGTCCCTTCATCGACGAAGAAGCGGAAGCACCTGTGATGAGATGA
>G1243 Amino Acid Sequence (domain in AA coordinates: 216-609)
MARNNSNDEAFSSEEEERVKNDEEEDDEEELAVARSSGSDDEVAADESPVSDGEAAP
VEDDYEDDEDEEKAEISKREKARLKEMQKLKKQKIQEMLSEQNASIDAMNNGKGRLLKY
LLQQTELFHFHFAKSDGSSSQKKAQGRGRHASKITEBEEDEEYLLKEEEDGLTSGNTRLLT
QPSCIQGMRDYQLAGLNLIRLYENGINGILADEMGLGKTLQTISSLAYLHEYRINGP
HMOVAPKSTLGNWMNEIRRFPCVLRVAVKFLGNPEERRHIREDLVAGKFDICVTSFEMAI
KEKTALRRFSWRYIIIDEAHRICKNENSLLSKTMRLFSTNYRLLITGTPLQNNLHELWALL

MDKEKSPAPPPSGGLPPPSGRYSAFSPNGSNFSAMKAESFFPLTPSGSNSDDANRFSHDI
SRMPDNPPKNLGHRRAHSEILTLPPDLLSFDSDLGVVGAADGPSFSDDTDEDLLYMYLDME
KFNSSATSTSQMGEPSEPTWRNELASTSNLQSTPGSSSERPRIHQHSQSMDGSTTIKPE
MLMSGNEDVSGVDSKKAISAAKLSELALIDPKRAKRIWANRQSAARSKERKMYIAELER
KVQTLQTEATSLSAQLTLLQRDTNGLGVENNELKLRVQTMEEQVHLQDALNDALKEEVQH
LKVLTQCGQPSNGTSMNYGSFGSNQQFYPPNNQSMHTILAAQQLQQLQIQSQKQQQQQQQH
QQQQQQQQQFHQQQLYLQQLQQRLQQQEQQSGASELRRPMPSPGQKESVTSFDPREPL
TKD*

ACTCAACTCGGAGGTATTGGTTTGAACCTAGCCGCTACTAATGGCAACAACCAAGCTCAC
CAGATCGGTTCCAGTTTGATGATGAGCGATCTAGGGTTTCTCCATGGACGAAATACTTCA
ACTCCGATGACGGGAAACATTTCATGAAAACAACAATAATAACAATGAAAACAACCTA
ATGGCATCCGTTGGATCTTTGAGCCCCCTTTGCTCTCTTCGATCCAACGACGGGGCTATAC
GCTTTCCAGAACGACGGTAATATCGGGAACAACGTTGGGATATCTGGTTCTTCTACTTCC
ATGGTTGATTCTAGGGTTTATCAGACGCCTCCGGTGAAGATGGAAGAACAACCTAATTTG
GCTAATTGTCTAGACCGGTCTCCGGTTTGACGTCTCCTGGGAATCAAACAATCAGTAC
TTTTGGCCTGGTTCCGATTTCTCGGGTCCTTCTAATGATCTCTTGTGA

>G1909 Amino Acid Sequence (conserved domain in AA coordinates:23-51)

MGGSMAERARQANIPPLAGPLKPCRDSSNTKFCYNNYNLTQPRHFCKGCRRYWTQGG
LRNVVPVGGGCRNRNKGKGNLKSSSSSSKQSSSVNAQSPSSGQLRTHQFPFSPILYNL
TQLGGIGLNLAAATNGNNQAHQIGSSLMMSDLGFLHGRNTSTPMTGNIHENNNNNNNNNL
MASVGSLSPPALFDPTTGLYAFQNDGNIGNNVGISGSSSTMVDSRVYQTPPVKMEBQPNL
ANLSRPVSGLTSPGNQTNQYFWPGSDFSGPSNDLL*

>G1663 (64..630)

TTCTCTCTGTGAATCCTTGTTTCATCGTCACTGAAATTAGTTTACAAAATCGACGAATTCG
GAGATGATTTTTTCAAGATGTGTGCAGAAATGAGTCCAACCTCAACGCTATAGCTTCCGAA
TCGCGTTCCCAAACGCAGTTCGGTGTTCGAAATCCTCCTCGAGCGGCGGCGGATGTATC
TCCGCCAGGACTAAAGACCGTCACACGAAGGTTAACGGACGAAGCCGTCGAGTTACGATG
CCGGCTCTCGCCGCGCTAGGATTTTCCAGTTAACGCGTGAGCTCGGTCACAAAACCTGAA
GGAGAAACCATCGAATGGCTTCTTAGTCAAGCTGAACCGTCGATTATTGCCGCCACTGGC
TACGGGACTAAGCTCATTTTGAATTTGGGTTGATGTTGCGGCGGACGATTCTCGTCGTCG
TCGTCGATGACGTCGCGCAAACGCAAACGCAAACGCCACAATCGCCGAGTTGTAGGTTG
GATCTTTGTGACCAATCGGAATTCAGTATCCGGTGAATGTTACAGTCATATGCCGTTT
ACAGCGATGCTTTTAGAGCCGATGACCACGACGGCGGAATCTGAGGTTGAGATCGCGGAG
GAGGAGGAACGTAGACGCCGTCACCATTAGTAAATATTAGGCTTTTGTATTAGAGTGTTAA
AATTAGGATTTTAAAGTTTAGGAGGTAACAGATAAGGATAATT

>G1663 Amino Acid Sequence (domain in AA coordinates: TBD)

MIFQNVCRNESNFNAIASESRSTQFVSKSSSSGGGCISARTKDRHTKVNRSRRVTMP
ALAAARIFQLTRELGHKTEGETIEWLLSQAEPSIIAATGYGKTLISNWVDVAADSSSSS
SMTSPQTQTQTPQSPSCRLDLCQPIGIQYPVNGYSHMPFTAMLLPMTTTAESEVEIAEE
EERRRRHH*

>G1231 (103..870)

CAAACCCAAATTCTCTCAGCGCCGGTCAAATACTTGTCTCTCTCTCTCTCTCTTTAC
TCTTGTCTTGTCTCCTTCGAAGCTGTTTGTCTGTGAAGAAAGATGGAAGCAGGTGGCGCG
TACAATCCACGCACTGTTGAAGAGGTGTTTAGGGATTTTAAGGGTCGTAGAGCTGGCATG
ATTAAGGCTTTAACCCTGATGTTTCAAGAGTCTTTTCCGACTTTGTGATCCCGAAAAGGAG
AACCTTTGCTTTTACGACATCCAAATGAGCACTGGGAAGTGAATTTGCCAGCTGAAGAG
GTTCTCTCTGAGCTCCAGAGCCTGTCTTGGGTATCAATTTTCCAGAGACGGGATGGCG
GAAAAGGATTGGTTGTCCCTTGTGCTGTCCACAGTGATGCTTGGCTTCTTGTGTTGCT
TTCTTTTTTGGAGCCAGGTTTGGATTTGACAAAGCTGATAGGAAGAGGCTTTTCAATATG
GTGAATGACCTCCCAACAATCTTTGAGGTTGTAGCTGGCACTGCTAAGAAACAAGGAAAA
GATAAGTCCTCTGTTTCCAACAACAGCAGCAACAGATCCAAATCAAGCTCCAAGCGAGGA
TCTGAATCCCGTGCCAAGTTCTCAAAGCCGAGCCCAAAGATGATGAGGAGGAGGAAGAG
GAAGGTGTGGAAGAGGAGGATGAGGATGAGCAAGGTGAACACAGTGTTGAGCATGTGGT
GAGAGCTATGCAGCTGATGAGTTCTGGATTTGCTGTGACCTCTGTGAGATGTGGTTTCAT
GGAAAGTGTGTTAAGATAACACCAGCAAGAGCTGAGCACATCAAGCAATACAAGTGCCCT
TCTTGACGAACAAAAGGGCTCGTTCCATAAATTTGTTGACCGCTCGCTTCTGTGTATCTA
CCTTTGCATATGATGATGAACAGCTTAACTGTTTGGTTTAGATCAGATTTGTATATGGA
TTTGGTAATTTAGGAAGACATTTTAGTTTTCATTGTTACATTTTGGCGATTGAAGGGA
TAACTCTTTGTTTAGGGGTAATGATCTTTTGCTCTGTTTTATGTTTGTATTAAACATTC
TTCAAACCTCAATCAAAAGTATTTTGGTTAGTCTTAAAA

>G1231 Amino Acid Sequence (domain in AA coordinates: TBD)

MEAGGAYNPRTVEFVFRDFKRRAGMIKALTDDVQEFFRLCDPEKENLCLYGHNPNEHWEV
NLPAEEVPPPELPEPVLGINFARDGMAEKDWLSLVAVHSDAWLLAVAFFFGARFGFDKADR
KRLFNMVNDLPTIFEVVAGTAKKQKDKSSVSNNSSNRSSSKRGSESRKFSKPEPKD
DEEEEEEGVEEEDDEQGETQCGACGESYAADEFWICCDLCEMWFHGKCVKITPARAEHI

KQYKCPSCSNKRARS*

>G227 (21..983)

GTACCGTCGACGATCCGGCGATGTCAAACCCGACCCGTAAGAATATGGAGAGGATTAAAG
GTCCATGGAGTCCAGAAGAAGATGATCTGTTGCAGAGGCTTGTTCAGAAACATGGTCCGA
GGAAGTGGTCTTTGATTAGCAAAATCAATCCCTGGACGTTCCGGCAAAATCTTGTCGTCTCC
GGTGGTGTAAACAGCTATCTCCGGAGGTAGAGCACCGTGCTTTTTTCGAGGAAGAAGACG
AGACGATTATTCGAGCTCACGCTCGGTTTGGTAACAAGTGGGCTACGATCTCTCGTCTTC
TCAATGGACGAACCGATAACGCTATCAAGAATCATTGAACTCGACGCTGAAGCGAAAT
GCAGCGTCGAAGGGCAAAGTTGTGATTTTGGTGGTAATGGAGGGTATGATGGTAATTTAG
GAGAAGAGCAACCGTTGAAACGTACGGCGAGTGGTGGTGGTGGTGTCTCGACTGGCTTGT
ATATGAGTCCCGGAAGTCCATCGGGATCTGACGTCAGCGAGCAATCTAGTGGTGGTGCAC
ACGTGTTTAAACCAACGGTTAGATCTGAGGTTACAGCGTCATCGTCTGGTGAAGATCCTC
CAACTTATCTTAGTTTGTCTCTTCTTGGACTGACGAGACGGTTCGAGTCAACGAGCCGG
TTCAACTTAACAGAATACGGTTATGGACGGTGGTTATACGGCGGAGCTGTTTCCGGTTA
GAAAGGAAGAGCAAGTGGAAGTAGAAGAAGAAGAAGCGAAGGGGATATCTGGTGGATTTCG
GTGGTGAAGTTTCATGACGGTGGTTCAGGAGATGATAAGGACGGAGGTGAGGAGTTACATGG
CGGATTTACAGCGAGGAAACGTGGTGGTAGTAGTTCTGGCGGCGAGGTGGCGGTTCGT
GTATGCCACAAAGTGTAACAGCCGTCGTGTTGGGTTTAGAGAGTTTATAGTGAACCAAA
TCGGAATTGGGAAGATGGAGTAGGCGGCC

>G227 Amino Acid Sequence (domain in AA coordinates: 13-112)
MSNPTRKNMERIKGPWSPEEDLLQRLVQKHGPRNWSLISKSIIPGRSGKSCRLRWCNQLS
PEVEHRAFQSQEEDETIIRAHARFGNKWATISRLLNGRTDNAIKNHWNSTLKRKCSVEGQS
CDFGGNGGYDNLGEEQPLKRTASGGGGVSTGLYMSPGSPSGSDVSEQSSGGAHVFKPTV
RSEVTASSSGEDPPTYLSLSLPWTDVTRVNEPVQLNQNTVMDGGYTAE LFPVRKEEQVE
VEEEEAKGISGGFGGEFMTVVQEMIRTEVRSYMDLQRGNVGGSSSGGGGGSCMPQSVN
SRRVGFREFIVNQIGIGKME*

>G1842 (219..809)

ACTATTACATGCCTCTTCTCGCTTCAAAACGGCACCGTTTCCACTTGTATTATTTTTC
TCTCTATCGTCTAACAAAAAAGTACTTGGGATTTTTTTTTCATTGTCTAGCCCA
AAAGAAGAAGATAGAAACGAAGAAAAAAGCAAAACACATTTTGGGTCCCCGGTGGTTAGG
ATCAAATTAGGGCACAAACCTTATCGGAGAAAGAAGCCATGGGAAGAAGAAAGTTCGAGA
TCAAGCGAATCGAGAACAAGCAGTCGACAAGTCACCTTCTCCAAACGACGCAAGGTC
TCATCGAAAAAGCTCGACAACCTTCAATTCTCTGTGAATCTCCATCGCTGTTGTGCGCG
TCTCCGGTTCGGGAAAACTCTACGACTCTGCCTCCGGTGACAACATGTCAAAGATCATTG
ATCGTTATGAAATACATCATGCTGATGAACCTAAAGCCTTAGATCTTGCAAAAAAATTC
GGAATTATCTTCCACACAAGGAGTTACTAGAAATAGTCCAAAGCAAGCTTGAAGAATCAA
ATGTCGATAATGTAAGTGTAGATTCTCTAATATCTATGGAGGAACAGCTCGAGACTGCTC
TGTCAGTAATTAGAGCTAAGAAGACAGAACTAATGATGGAGGATAGAAGTCACCTTCAAG
AAAGGGAGAAGTTGCTGATAGAAGAGAACCAGATTCTGGCTAGCCAGGTGGGGAGAAGA
CGTTTCTGGTTATAGAAGGTGACAGAGGAATGTACGGGAAAATGGCTCCGGCAACAAAG
TACCGGAGACTCTTTCGCTGCTCAAGTAATCACCATCATCAACGGCTGAGCTTTCACCAT
AACTTACTCACAGCTGATTGAGAAGCTTTTACAAAATTGTAAATTATAAAAAGCTGCA
TAATAATCTCAACCTTTTATCTTCTCGCGCAATGTGGAAATAAAGGTAAACAAAAC
GAAGCTCTTTTCTTTATGCGAAAGAAATTGTAAACTAAGATAAAGCTACCGATCTTTGT
TGTACCTTAGTAGACAAATATCAGAGTTCTTGTGCTTGT

>G1842 Amino Acid Sequence (domain in AA coordinates: 2-57)
MGRRKVEIKRIENKSSRQVTFSKRRKGLIEKARQLSILCESSIAVVAVSGSGKLYDSASG
DNMSKIIDRYEIHHADELKALDLAEKIRNYLPHKELLEIVQSKLEESNVDNVSDSLISM
EEQLETALSVIRAKKTELMEDMKSLQEREKLLIEENQILASQVGKKTFLVIEGDRGMSR
ENGSGNKVPETLSLLK*

>G1505 (1..681)

ATGGATGATATAGCGGAACCTGAATGGTTATCAAATTCGTAGATGATTCTTCTTTCACG
CCGTATTCTGCTCCGACGAATAAACCGGTTTGGTTAACCGGAAATCGGAGACATCTGTA
CAACCGGTTAAAGAGGAGACCTGCTTCAAATCCCAACATCCGGCCGTCAAACAGACCC
AAACGACCCAGAACCGGAGTCAGAGTCTGGTCTCATGGTTCGCACTCGTTAACCGACTCA
TCTTCAAGCTCTACAAACATCTTCTGCTCCTCTCCTCGTCTTCAAGCCCTCTATGGCTC
GCCAGCGGTGAGTTTCTTGTAGAGCAATGACTAAAACACAAAAGAAGAAGAAAGTTTGG

AAAAACGCTGGTCAGACGCAAACGCAAACGACGACGACGCGGCAGTGTGGTCATTGT
GGAGTTCAGAAAACGCCGAGTGAGAGCAGGACCATTAGGAGCGAAGACGTTGTGTAAT
GCGTGTGGTGTGCGTTACAAATCGGGTCTGTTACTACCCGAATATAGACCCGCTTGTAGC
CCAACATTTTCGAGTGAGCTTCACTCAAACCACCACAGTAAAGTCATTGAGATGCGTAGG
AAGAAAGAGACTTCTGACGGTGCTGAAGAAACCGGTTTGAACCAGCCGGTTCAGACGGTT
CAGGTTGTCTCGAGTTTTTGA

>G1505 Amino Acid Sequence (domain in AA coordinates: TBD)
MDDIAELEWLSNFDSSFTTPYSAPTNKPVWLTGNRRHLVQPVKEETCFKSOHPAVKTRP
KRARTGVRVWSHGSQSLTDSSTSTSSSSPRPSSPLWLASQFLDEPMTKTQKKKKVW
KNAGQTQTQTQTQRCGHCQVQKTPQWRAGPLGAKTLCNACGVRYKSGRLLPEYRPACS
PTFSSELHSHNHSKVIEMRRKETS DGAETGLNQPVQTVQVVSSF*

>G657 (1..2331)

ATGAAGCGTGAGATGAAAGCACCTACTACTCCACTAGAGAGTCTCCAAGGTGACCTCAAA
GGAAAACAAGGGAGGACATCTGGCCCTGCTAGACGATCTACCAAAGGACAATGGACACCT
GAAGAGGACGAAGTCTTGTGTAAAGCTGTTGAGCGTTTTCAAGGAAAGAACTGGAAGAAG
ATAGCTGAATGTTTTAAGGATCGGACTGATGTTTCAAGTGTCTTCATAGATGGCAAAGGTC
TTGAACCCAGAGCTTGTGAAAGGACCGTGGTCAAAAGAGGAGGATAACACAATAATTGAC
CTGGTTGAAAAATATGGGGCAAAGAAATGGTCTACTATATCTCAGCATTACCTGGGCGC
ATAGGAAAGCAATGTAGGGAAAGGTGGCATAACCATCTTAACCTGGGATTAATAAAAAT
GCATGGACTCAGGAAGAGGAAGTACTCTTATTCGTGCGCATCAAATTTATGGGAATAAA
TGGGCAGAGCTTATGAAATTTTTGCGAGGAAGGTCAAGTAATTCGATAAAAAATCATTGG
AACAGCTCAGTTAAGAAGAAGTTGGATTCTACTATGCATCAGGTCTTTTAGATCAGTGT
CAAAGCTCGCCATTAATTGCCCTTCAGAACAAATCTATCGCTTCATCTTCCTCGTGGATG
CACAGCAATGGAGATGAAGGTAGTTCAAGGCCAGGGGTGATGCTGAGGAATCAGAATGC
AGCCAAGCTTCAACTGTTTTCTCACAATCAACCAACGATTTACAAGATGAAGTTCAACGT
GGAAATGAGGAATATTACATGCCTGAATTTTCAATTCAGGAACGGAGCAGCAAATCTCAAAC
GCTGCATCTCATGCAGAACCGTACTACCTTCTTTAAAGATGTCAAATTTGTGTCCCC
GAAATTTCTTGTAACAGAAATGTTTCAAGAAGTTTCAAGATCTTAATTTGTCTCACGAG
CTAAGAATACCAACAGCTACGGAGGATCAATTGCCGGGTGATCTAATGATGCTAAACAG
GACCGTGGTCTAGAGTTATTGACCCATAACATGGACAACGGTGGAAAAAACCAAGCACTT
CAACAAGATTTTCAAAGTTTCAAGATTAAGTGATCAACCTTTTTTGTCAAACCTCGGAC
ACAGATCCAGAAGCTCAAACCTTTGATCAGGATGAGGAGTGTGTAGGGTTCTTTTCCA
GATAACATGAAAGATAGCAGTACATCTTCTGGTGAGCAAGGTGCGGAATATGGTTGACCT
CAAAACGGCAAAGGATCTCTTTGTTCTCAGGCTGCAGAAACCATGCTCATGAAACTGGA
AAAGTTCCAGCTTTACCGTGGCATCTTCAAGTTCTGAGGGCCTGGCGGGTCATAATTGT
GTCCCTTTGTTGGATTGAGCTTTGAAGGACTCACTTTTACCCGTAATGATTCCAACGCT
CCTATAAGAAGTTGTGCGCTTTTTTGGAGCTACCGAATTAGAATGTAAGACTGATACAAAT
GACGGTTTTTCGATACCTTACGGACATGTAACCTCCCATGGCAATGATGATAATGGTGGT
TTCCAGAACAAACAGGGGCTGTATATATCCCAAGGATCTTTGAAGCTAGTACCTTTG
AATAGTTTTTCTCTCTCTAGAGTGAACAAGATTTATTTCTATTGACGATAAGCCG
GCTGAAAAAGACAAAGGAGCTCTTTGTATGAACCTCCACGTTTTCCAAGTGCAGATATT
CCTTTCTTCAGCTGTGATCTTGTACCATCAAATAGTGACTTACGGCAAGAGTACAGTCCC
TTTGGTATCCGTCAGTTGATGATTTCTCAATGAATTGTACAACTCCGTTAAGGTTATGG
GATTCACCGTGTACGATAGGAGCCCTGATGTGCTTAATGATACTGCCAAAAGTTTT
AGTGGTGACCATCCATCTTAAGAAGCGGCATCGAGACTTGCTTTACCTGTGCTTGAT
AGAAGAAAAGACAAAAAGCTTAAAGGGCTGCGACTTCTCTTGGCTAATGATTTTTCG
CGCTTAGATGTAATGCTTGATGAAGGAGATGATTGCATGACCTCTCGTCCGTCAGAGTCT
CCTGAAGATAAAAAATATGTGCTCCCTTCCATAGCCAGAGATAACAGAAATGTGCA
TCAGCTCGGTTATATCAAGAAATGATTCCGATAGATGAGGAACCAAAGGAAACCTTAGAA
TCAGGTGGAGTGACTTCTATGCAAAATGAAATGGATGTAATGACGGTGGTGTTCAGCT
AAAAATGTAAGTCCGCTCTTTGTCTTCATATTATCTGGTATCAGTTATAA

>G657 Amino Acid Sequence (domain in AA coordinates: TBD)
MKREMKAPTTPLBSLQGDLDKKGQRTSGPARRSTKGQWTPPEDEVLCMAVERFQGNWKK
IAECFKDRDTPVQCLHRWQKVLNPELVKGPWSKEEDNTIIDLVEKYGPKKWSTISQHLPGR
IGKQCRERWHLNPNPINKNAWTQEEELTLIRAHQIYGNKWAELMKFLPGRSDNSIKNHW
NSSVKKKLDSSYASGLLDQCQSSPLIALQNKSIASSSSWMHNSNGDEGSSRPVDAEES
SQASTVFSQSTNDLQDEVQRGNEEYMPFHSQTEQQISNAASHAEPYYPSPKDVKIVVP

EISCETECSSKKFQNLNCSHELRTTTATEDQLPGVSNDKQDRGLELLTHNMDNGGKNQAL
QQDFQSSVRLSDQPFLLNSDTPDPAQTLITDEECCRVLPDNMKDSSTSSGEQGRNMVDP
QNGKGLCSQAAETHAHETGKVPALPWHPPSSSEGLAGHNCVPLLDSDLKDSLLPRNDSNA
PIQGCRFLFGATELECKTDTNDGFIDTYGHVTSHGNDNGGFPEQQGLSYIPKDSLKLVLPL
NSFSSPSRVNKIYFPIDDKPAEKDKGALCYEPPRFPSADIPFFSCDLVPSNSDLRQEYSP
FGIRQLMISSMNCTTPLRLWDSPOCHDRSPDVMLNDTAKSFSGAPSILKKRHRDLLSPVLD
RRKDKKLKRAATSSLANDFSRLDVMLDEGDDCMTSRPSESPEDKNICASPSIARDNRNCA
SARLYQEMIPIDEEPKETLESGGVTSMQNENGENDGGASAKNVSPSLSLHIIWYQL*

>G1959 (141..1028)

CGTCGACTGTCCATAAATCCGGAGCCTGACCCGACGTTTGACCCGGATCCGAACTCCCA
CAATCTCCATACCACCCAAATTCATCTCCCTAAAGCTTTCTCTCACTTTCCCGGAAAA
TCGGCGACCAAAATGGAAAATGTACTCAGCGATTTCGCTCGCTTCCACTCGATGGTGGAC
ACGTTGGTGGTGACTACCATGGACCTTTGACGGAACCAATCTTCCCGGTGACGCTTGTT
TGGTTTTTAACGACTGACCCCTAAACCTCGTCTCCGGTGGACAACTGAGCTTCATGAGAGAT
TCGTTGACGCCGTTACTCAGCTCGGTGGTCTGACAAAGCGACTCCCAAACTATTATGA
GAACAATGGGAGTGAAGGGTCTCACTCTTACCACCTCAAATCACATCTTCAGAAATTC
GCCTAGGGAGGCAAGCTGGCAAAGAATCAACTGAGAATCTTAAAGATGCTTCTGTGTAG
GGGAGAGTCAGGACACAGGTTTCTTTCGACATCATCAATGAGAATGGCGCAGCAGGAGC
AGAACGAGGGTTACCAAGTCACCGAAGCTCTACGTGCTCAGATGGAAGTCCAAAGAAGAC
TACACGATCAATTGGAGGTGCAACGGAGGCTCCAGCTGAGGATAGAGGCACAAGGAAAAT
ACCTGCAATCGATTCTTGAAAAAGCTTGCAAGGCCTTTGACGAGCAAGCTGCTACTTTTG
CTGGACTTGAGGGTCTTAGGGAAGAGCTATCAGAGCTAGCCATCAAAGTCTCCAATAGCT
CTCAAGGAACATCAGTCCCGTACTTTCGATGCAACAAAGATGATGATGATGCCATCGTTGT
CAGAGCTTGAGTAGCAATAGACAACAAAAACAACATCACAACTGTTTCACTAGAAA
GCTCTCTGACTTCCATCACATGGGAGCTCTATATCTGCTGCATCAATGAAGAAGCGTC
AACGTGGAGACAAATTTGGGCGTAGGGTATGAATCAGGCTGGATTATGCCTAGTAGACCA
TTGGATAAAGTTTAGGAGAGGGAAAAAGTTTATTATGGGAAAGGTAGAGATAAGATTAA
CTGTTCTTTACTTGTCTTGAGGGGCTGCGGCCGCT

>G1959 Amino Acid Sequence (conserved domain in AA coordinates:46-97)

MYSAIRSLPLDGGHVGVDYHGPLDGTNLPGDAQLVLTDPKPRLRWTTTELHERFVDAVTQ
LGGPDKATPKTIMRTMGVKGLTLYHLKSHLQKFRLLGRQAGKESTENSKDASCVGESQDTG
SSSTSSMRMAQQEQNEGYQVTEALRAQMEVQRRLHDQLEVQRRLQLRIEAQGYLQSILE
KACKAFDEQAATFAGLEAAREELSELAIKVSNSSQGTSPYPFDATKMMMPSLSELAVAL
DNKNNITTNCSVESLTSITHGSSISAASMKKRQRGDNLGVGYESGWIMPSSTIG*

>G2180 (1..1440)

ATGGCTCCTGTCTCGTTACCTCCAGGTTTCCGATTCCATCCAACAGACGAGGAACATAATT
ACTTACTACTCTAAAAAGAAAGATCAACGGTCTAGAAATCGAACTTGAAGTTATCGCTGAA
GTTGATCTTTTACAAGTGTGAGCCATGGGACTTACCAGGGAAGTCCTTGCTTCCGAGCAAA
GACCAAGAATGGTACTTCTTTCAGCCCACGAGACCGGAAGTATCCCAACGGCTCAAGGACA
AACCAGGCAACTAAAGGCGGTTATTGGAAGGCTACAGGTAAAGACCGCCGAGTTAGTTGG
AGAGACCGAGCCATAGGAACCAAGAAGACATTGGTTTTACTACCGTGGGCGCGGCCCAT
GGCATAAGAAGTGGTTGGGTGATGCACGAATATCGACTTGATGAAACAGAATGTGAGCCT
TCTGCATACGGCATGCAGGACGCATATGCACTTTGTCTGTGTGTTCAAAAAGATTGTTATT
GAAGCTAAGCCAAGAGATCAACATCGGTCATATGTCCACGCGATGTGCAATGTGAGTGGT
AATTGCTCATCGAGTTTGTGACACTTGTTCGGATCTCGAAATCAGTTCAACTACTCATCAA
GTTCAAAAACACATTCCAACCGCGATTGCGCAACGAGCGATTAACTCCAACGCAATCAGC
AACCAGGATTGGTCACAATACTACGGTCTTCTTATAGACCGTTCCTACTCCATATAAG
GTTAACACAGAGATGGAATGTTCAATGTTACAACACAATATATATCTACCACCGTTGCGT
GTAGAGAACTCTGCGTTTAGTGATTCCGATTCTTTCACGAGTATGACTCACAACAACGAC
CATGGCGTTTTTCGATGACTTTACTTTTGCTGCAAGTAACCTCAACCACAATAATAGCGTT
GGTGATCAAGTGATCCACGTTGGCAATTATGATGAACAATTAATAACATCTAACCCTCAT
ATGAACCAGACTGGTTATATAAAAGAGCAGAAGATCAGATCGAGTTTGGATAATACTGAC
GAAGTACAGGATTTTATGGTAACAATACCAATGACAACATAGATATCGATGATTTTCTC
TCGTTTGATATATATAACGAGGACAACGTGAATCAAATAGAAGATAATGAAGACGTGAAT
ACAAATGAAACCTTGATTTCATCGGGATTTCGAGGTGGTTGAAGAAGAACTAGATTTAAC
AACCAAATGCTCATCTCGACATATCAAACGACAAAGATTCTATATACCAAGTCGTACCT
TGTCACACGTTGAAAGTTCAGTCAATCTATTAGTCACAATGTGGAAGAGAGAACATTG

TTCATTGAAGAGGACAAAAGATTCTTGGTTACAAAGAGCTGAGAAGATCACGAAGACAAAA
CTAACACTTTTTAGTTTAATGGCTCAGCAATACTACAAATGTCTTGCTATTTTTTCTGA
>G2180 Amino Acid Sequence (conserved domain in AA coordinates:7-156)
MAPVSLPPGFRFHPTDEELITYYLKRKINGLEIELEVIAEVDLYKCEPWLPGKSLPSK
DQEWYFFSPDRKYPNGSRTNRATKGGYWKATGKDRRVSWRDRAIGTKKTLVYYRGRAPH
GIRTGWVMHEYRLDETECEPSAYGMQDAYALCRVFKKIVIEAKPRDQHRSYVHAMSNSVSG
NCSSSFDTCSDLISSTTHQVQNTFQPRFGNERFNSNAISNEDWSQYYGSSYRPFPTPYK
VNTEIECSMLQHNIYLPPLRVENSASFSDSFFTSMTNNDHGVFDDFTFAASNSNHNNSV
GDQVIHVGNVDEQLITSNRHMNQTYIKEQKIRSSLDNTDEDPGFHGNNNTNDNIDIDDFL
SFDIYNEDNVNQIEDNEDVNTNETLDSSGFVVEEETRFNNQMLISTYQTTKILYHQVVP
CHTLKVHVNPISHNVEERTLFIEEDKDSWLQRAEKITKTKLTLFSLMAQQYYKCLAIFP*
>G1817 (1..1308)
ATGAAGGACGCGAGAGAAGCGAGAGGTGATTGCATCATCATCATTTACAAAGAAAGAGAAAC
AGAGGAAGAAGACTAAGGAAAAGAAGAAGAAGAACGAGAAGCGAGTACTAATGGTTCCA
TCATCATTACCAAACGACGTGCTAGAGGAGATCTTTTTAAGATTTCCGGTTAAAGCCCTA
ATCCGACTCAAGTCTCTCTCGAAACAATGGAGATCGACGATCGAATCTCGCAGTTTTGAA
GAGAGACACTTGACGATCGCTAAGAAAGCCTTCGTGGATCATCCAAGGTCATGCTCGTA
GGAGAAGAAGATCCCATAGAGGAACCGGATTCGTCCAGACACTGACATTGGTTTTAGG
TTATTCTGCTTGGGAATCGGCTTCTCTCTATCCTTTACTCGTCTCAATTTCCCTCAAGGG
TTCTTCAACTGGATCTACATATCTGAAAGCTGTGATGGCCTTTTCTGCATCCATTTCCCA
AATACATATCCGTATATGTAGTGAATCCGGCTACACGGTGGCTCCGCTACTTCTCTCCG
GCAGGGTTTTAGATTTTGTATCCACAAGTTTAACCCCACTGAACGTGAGTGAATGTAGTG
ATGAAATCAATCTTTCATCTAGCATTCGTGAAGGCCACCGATTACAAATTAGTGTGGTTG
TACAATTGTGATAAGTACATTGTTGATGCGTTCGAGTCCAAACGTGGGAGTCACAAAGTGC
GAGATTTTTGACTTTAGGAAAAATGCTTGGAGGTACTTGGCTTGCACTCCAAGTCATCAG
ATATTCTATTACCAAAAGCCAGCATCTGCAAACGGGTCGGTTTATTGGTTTACAGAACCA
TATAATGAAAGAAATCGAAGTAGTGGCTTTTGATATTACAGACCGAAACATTCCGGTTGCTG
CCTAAGATTAATCCGGCTATTGCTGGTTCAGATCCTCACCATTATGACATGTGCACTCTG
GATAATAGTTTTGTGTGTCGAAAAGGAGAGAAAGATACTATGATCCAAGATATTTGGAGG
TTGAAACCATCAGAAGACACATGGGAAAAGATTTTTAGCATAGACTTGGTTTTCTGTCTCT
TCTTCTCGGACTGAGAAGCGTGATCAATTTGATTGGAGCAAGAAGGATAGGGTTGAGCCA
GCCACACCCGTCGCGGTTTTGTAAGAATAAGAAGATCCTTCTCTCACATCGCTATTCCCGA
GGTTTTGGTAAAGTACGATCCCCTAACAAATCTATCGATTTTTTTTCCGGACATCCTACC
GCTTACAGAAAAGTTATTTATTTTCAAAGTTTGATATCTCATCTATAA
>G1817 Amino Acid Sequence (conserved domain in AA coordinates:47-331)
MKDAEKREVIASSSLQQRKRNRRRLRKRNRNEKRVLMVPSSLPNDVLEEIFLRFVVKAL
IRLSLSKQWRSTIESRSFEERHLTIAKKAFFVDHPKVMLVGEEDPIRGTGIRPDTDIGFR
LFCLESASLLSFTRLNFPQGFNWIIYISESCDGLFCIHSPKSHSVYVVPATRWLRLLPP
AGFQILIHKNPTEREWNVVMKSIHFLAFVKATDYKLVWLYNCDKYIVDASSPNVGVTKC
BIFDFRKNAWRYLACTPSHQIFYQKPAANGSVYWFTEPYNERIEVVAFDIQETETFRLL
PKINPAIAGSDPHHIDMCTLDNSLCMSKREKDTMIQDIWRLKPS EDTWEKIFSIDLVSCP
SSRTEKRDQFDWSKKDRVEPATPVAVCKNKKILLSHRYSRGLVKYDPLTKSIDFFSGHPT
AYRKVIYFQSLISHL*
>G1649 (61..1311)
ATTCACAAAAACCGGAAAAAAGACAAGTAAAGAAAGCTTTGTTCAAGTTTACTTCA
ATGGAAGCAAAACCTTAGCATCATCATCTGAACCAACATGATTTCTCCATCATCA
AACATTAAACCAAAATTAAGATGAAGATTATATGGAGCTGGTGTGTGAAAATGGGCAG
ATTCTTGCAAGATTCGAAGACCAAGAACAACGGTTCCTTTCAAAGCAACGTAGGCAA
TCTCTCCTGGATTTGTATGAGACCGAGTACAGCGAGGGTTTCAAGAAAAACATCAAGATT
CTTGGAGACACACAAGTTGTTCCGGTGAGTCAGTCTAAGCCACAACAAGATAAAGAAACC
AATGAACAAATGAACAACAATAAGAAGAGCTAAAGTCCTCCAAAATCGAATTTGAGAGA
AATGTTTCGAAAAGCAACAATGTGTTGAATCATCAACATTAATTGATGTTTCTGCTAAA
GGTCCAAAGAAATGTTGAAGTTACTACAGCTCCTCCTGATGAGCAATCTGCAGCTGTGGT
AGATCCACGGAATTGTATTTTGCTTCTCATCGAAGTTTCTCGAGGAACCTCGAGAGAT
CTAAGTTGTTGTTCTTTAAAGAGGAAGTATGGAGATATTGAAGAAGAAGAATCAACCTAT
TTAAGTAATAATTCAGATGATGAATCAGATGATGCGAAGACACAAGTTCATGCGAGAAC
AGAAAGCCGGTGACTAAAAGAAAACGAAGCACAGAAGTCCATAAGTTATATGAAAGAAAA

CGAAGAGATGAATTCAACAAGAAAATGCGTGCTTTGCAGGACCTACTACCAAATTGTTAC
AAGGATGATAAGGCTTCATTGTTGGATGAGGCTATCAAATATATGCGGACCCTTCAACTT
CAAGTTCAGATGATGAGTATGGGAAATGGATTAATAAGACCACCTACGATGTTGCCAATG
GGTCATTACTCTCCCATGGGTCTAGGAATGCATATGGGTGCAGCAGCAACACCAACATCA
ATACCCGAATTCCTGCCTATGAATGTTCAAGCAACCGGTTTCCGGGGATGAACAATGCA
CCACCACAAATGCTAAGCTTTCTTAATCACCCAAGTGGACTAATTCCAAACACTCCTATC
TTTTCTCCATTGGAAAATTGCTCTCAGCCATTCTGTTGGTGCCTTCGTGTGTTTCTCAGACT
CAGGCTACTTCTTTTACTCAATTCCCAAAGTCTGCGTCCGCTCAAACCTAGAAGATGCA
ATGCAATATAGAGGAAGCAACGGTTTTAGTTATTATCGCTCGCCAACTAATGATTGTA
GAAAGTTGATGTTTCTCCAACCTAACTTTAAGCAAAAAAATGATCGTCTACTCT
GTGTTGTTAGTCTATGGGCTTTTGGGCCCTTGATTCTTGGAACGATTTGAACCTAATTCCA
ACTATTTTCAAAGTGGATGTACAAAGTAAAA

>G1649 Amino Acid Sequence (conserved domain in AA coordinates:225-295)

MEAKPLASSSEPNNMISPSSNIKPKLKDEDYMEVLCENGQILAKIRRPKNNGSFQKQRRQ
SLLDLYETEYSEGFKNIKILGDTQVVPVSQSKPQDKETNEQMNNKKKLKSSKIEFER
NVSKSNKCVESSTLIDVSAKGPKNVEVTTAPPDEQSAAVGRSTELYFASSSKFSRGTSRD
LSCCSLKRKYGDIEEEESTYLSNNSDDSDDAKTQVHARTRKPVTKRKRSTEVHKLKERK
RRDEFNKKMRALQDLLPNCYKDDKASLLDEAIKYMRTLQLQVQMMSMGNGLIRPPTMLPM
GHYSPMGLGMHMGAAATPTSIPQFLPMNVQATGFPGMNAPPQMLSFLNHPGLIPNTPPI
FSPLENCSPFVVPSCVSQTQATSFTQFPKSASASNLEDAMQYRGSNGFSYRSPN*

>G2131 (69..1010)

GTCTCTCATTTCATAATTCATTTTCAGGATTGTCTCTCAATCTTTTATTCTTCTCATT
CACCGTAATGGCAAAAAGTCTCTGGGAGGAGCAAGAAAACAATCGTTGACGATGAAATCA
GCGATAAAACAGCGTCTGCGTCTGAGTCTGCGTCCATTGCCCTAACATCCAAACGCAAAAC
GTAAGTCGCCGCCCTCGAAACGCTCCTCTTCAACGCAGCTCCCTTACAGAGGCGTCACAA
GGCATAGATGGACTGGGAGATACGAAGCGCATTTGTGGGATAAGAACAGCTGGAACGATA
CACAGACCAAGAAAGGACGTCAAGTTTATCTAGGGGCTTACGACGAAGAAGAAGCAGCAG
CACGTGCCTACGACTTAGCAGCATTGAAGTACTGGGGACGAGACACACTCTTGAACITCC
CTTTGCCGAGTTATGACGAAGACGTCAAAGAAATGGAAGGCCAATCCAAGGAAGAGTATA
TTGGATCATTGACAGAAAGTAGTGGATTCTCTCGCGGTGTATCAAAATACAGAGGCG
TTGCAAGGCATCACCAATAATGGGAGATGGGAAGCTAGAATTGGAAGGGTGTGTTGGTAATA
AATATCTATATCTTGGAACATACGCCACGCAAGAAGAAGCAGCAATCGCCTACGACATCG
CGGCAATAGAGTACCGTGGACTTAACGCCGTTACCAATTTGACGTCAGCCGTTATCTAA
ACCCTAACGCCGCCCGCGGATAAAGCCGATTCCGATTCTAAGCCCATTCGAAGCCCTAGTC
GCCAGCCCGAATCGTCCGATGATAACAAATCTCCGAAATCAGAGGAAGTAATCGAACCAT
CTACATCGCCGGAAGTGATTCCAACCTCGCCGGAGCTTCCCGACGATATCCAGACGTATT
TTGGGTGTCAAGATTCGGCAAGTTAGCGACTGAGGAAGACGTAATATTGATTGTTTCA
ATTCTTATATAAATCTGGCTTCTATAACGAGTTTGATTATGGACCTTAATCGTATTTTC
TACAAGTTTTGTTTTGATTATCTACACAATACATCAATATATTCT

>G2131 Amino Acid Sequence (conserved domain in AA coordinates:50-186, 112-183)

MAKVSGRSKKTIVDDEISDKTASASESASIALTSKRKRKSPPRNAPLQRSSPYRGVTRHR
WTGRYEAHLWDKNSWNTQTKKGRQVYLGAYDEEEAAARAYDLAALKYWRDTRLNFPPLP
SYDEDVKEMEGQSKEEYIGSLRRKSSGFSRGSVSKYRGVARHHHNGRWEARIGRVFGNKYL
YLGTYATQEEAAIAYDIAAIEYRGLNAVTNFDVSRYLNPNAADKADSDSKPIRSPSREP
ESSDDNKSPPKSEEVIEPSTSPFVPTRRSFDDIQTYFGCQDSGKLATEEDVIFDCFNSY
INPGFYNEFDYGP*

>G215 (1..1110)

ATGACTCGTCCGGTGTTCGCATTGTAGCAACAATGGGCACAATTCACGCACGTGTCCAACG
CGTGGGTCTGGTTCCTCCTCCGCCGTGAAGTTATTTGGTGTGAGGTTAACGGATGGCTCG
ATTATTAAGAGAGTGCAGTATGGGTAATCTCTCGGCATTGGCTGTTGCCGCGCGCGCG
GCAACGCACCACCGTTTATCTCCGTCTCTCTCTGGCGACGTCAAATCTTAATGATTCTG
CCGTTATCGGATCATGCCGATACTCTAATTTGCATCATAATGAAGGGTATTTATCTGAT
GATCCTGTCTAGTCTGGGTCTAGTCACCGTCGTGGTGAGAGGAAGAGAGGTGTTTCT
TGGACTGCAAGAGGAACATAGACTATTCTTAGTCGGTCTTCAGAACTCGGGAAAGGAGAT
TGGCGCGGATTTTCGAGAAAATATGTAACGTCAGAACTCTACACAAGTGGCTAGTCAT
GCTCAAAAGTATTTTATTCGACATACTAGTTCAAGCCGAGGAAAAGACGGTCTAGCCTC
TTCGACATGGTTACAGATGAGATGGTAACCGATTATCGCCAACACAGGAAGAGCAGACC

TTAAACGGTTCCTCTCCAAGCAAGGAACCTGAAAAGAAAAGCTACCTTCCTTCACTTGAG
CTCTCACTCAATAATACCACAGAAGCTGAAGAGGTCGTAGCCACGGCGCCACGACAGGAA
AAATCTCAAGAAGCTATAGAACCATCAAATGGTGTTCACCAATGCTAGTCCCAGGTGGC
TTCTTTCTCCTTGTTCCTTCCAGTGACTTACACGATTGGCTCCCTGCGTCACTTCACGGA
ACAGAACATGCCTTAAACGCTGAGACTTCTTCTCAGCAGCATCAGGTCCTAAAACCAAAA
CCTGGATTGCTAAAGAACGTGTGAACATGGACGAGTTGGTGGTATGTCTCAGCTTAGC
ATAGGAATGGCGACAAGACACGAAACCGAAACTTCCCTTCCCGCTATCTTTGAGACTA
GAGCCCTCAAGGCCATCAGCGTTTCACTCGAATGGCTCGGTAAATGGTGCAGATTGAGT
AAAGGCAACAGCGCGATTCAAGGCTATCTAA

>G215 Amino Acid Sequence (domain in AA coordinates: TBD)
MTRRCSHCSNNGHNSRTCPTRGSGSSSAVKLFGVRLTDGSIKKSSASMGNLSALAVAAAA
ATHHRLSPSSPLATSNLNDSPLDHARYSNLHHNEGYSDDPAHSGSGSSHRRGERKRGVP
WTEEEHRLFLVLGLQKLKGKGDWRGISRNYVTSRPTQVASHAQKYFIRHTSSSRKRSSSL
FDMVTDEMVTDSPTQEEQTLNGSSPSKEPEKKSYPLESLNNTTEAEVAVATAPRQE
KSQEAIEPSNGVSPMLVPGGFFPFCFPVYTIWLPASLHGTEHALNAETSSQHQVLPKPK
PGFAKERVNMDELVGMSQLSIGMATRHETETSPSPSLRLLEPSRPSAFHSNGSVNGADLS
KGNSAIQAI*

>G1508 (1..420)

ATGCTAGATCACAGTAAAAGGCTCTTATTGGTTGATTGAGAAACCATGAAAACAAGAGCT
GAAGATATGATCGAACAGAACAACTAGTGTAAACGACAAGAAGAGACTTGTGCTGAT
TGTGGAACAGTAAACTCCTCTTTGGCGTGGTGGTCTGTTGGTCCAAAGTCGTTGTGT
AACGCGTGTGGGATCAGAAACAGAAAGAAGAGAAGAGGAGGAACAGAAAGATAATAAGAAA
TTAAAGAAATCGAGTTCTGGCGGCGGAAACCGTAAATTTGGTGAATCGTTAAACAGAGT
TTGATGGATTTGGGGATAAGGAAGAGATCAACGGTGGAGAAGCAACGACAGAAGCTTGGT
GAAGAAGAACAGCCGCTGTGTTACTCATGGCTCTTTCTTATGGCTCTGTTTACGCTTAG

>G1508 Amino Acid Sequence (domain in AA coordinates: 38-63)
MLDHSEKVLVLDSETMKTRAEDMIEQNNTSVNDKKKTCADCGTSKTPLWRGGPVGPKSLC
NACGIRNRKKRRGGTENDKKLKKSSSGGNRKFGBSLKQSLMDLGIRKRSTVEKQRQKLK
EEEQAQAVLLMALSYGVSVA*

>G2110 (36..1622)

GAGAGCTAATAAAAAATTTATCAAAGAAGACTAATATGGAGAAGGACGATTTCTTGAGGA
GTGTCATGGAAGAGAAGAAAGCCATGATGAGATGAGAAAACCTTGATTCATCTCACGATG
ATTCTCATCAAGAACACGACCATATTATAAGATCCAAGTTGGACTCAACTAAAGTCGAAA
TGGATGAGGCTAAAGAGGAAAATCGAAGACTAAAGTCATCATTGAGTAAATCAAGAAAG
ATTTTGACATCCTTCAAACACAATACAACCAATTAATGGCCAAACATAACGAACCAACCA
AGTTCCAATCAAAAGGGCATCATCAAGACAAAGGCGAAGATGAAGACAGAGAAAAAGTTA
ACGAACGTGAAGAACTTGTCTCGTTGAGCCTAGGCAGACGGTTAAATTGAGAGTTCCAA
GTGGTTCGAATAAAGAAGAAAAAATAAAGATGTTGAAGAAGCGGAAGGTGACAGAAATT
ATGATGATAATGAAAAAGCAGTATTCAAGGGTTGAGTATGGGGATTGAATACAAGGCTT
TGAGTAATCCTAATGAGAAGTTAGAGATTGATCATAATCAAGAAACCATGTCGTTGGAGA
TTAGTAACAATAAAGATCAGATCACAAAATAGTTTTGGGTTAAGAATGATGGAGATG
ATCATGAAGATGAAGATGAGATTTGCCTCAAAACCTTGTTAAGAAAACTAGGGTTTCGG
TGAGATCAAGATGTGAGACACCAACGATGAACGACGGATGTCAATGGAGGAAATATGGCC
AGAAAATAGCTAAAGGCAATCCATGTCCCGAGCTTACTATCGTTGCACCATTCGAGCTT
CTTGTCAGTAAGAAAAACAGGTGCAAGATGTTGAGAAGATATGCTATACCTTATCTCAA
CGTACGAAGGAACACATAACCATCCACTTCCCATGTCAGCAACTGCCATGGCCTCTGCCA
CTTCCGCTGCCGCTCCATGCTTCTCTCGGCGCCTCCTCCTCATCCGCCGCGAGCTG
ATCTTCATGGCCTTAACTTCTCTCTTCCGGCAACAACATCACTCCAAAACCTAAAACTC
ATTTCTCCTCAATCCCCTTCTTCTTCTGCGCATCCGACCGTCACTCTCGACCTCACAACCT
CCTCCTCGTCGAGCAACCGTTCTTATCAATGCTCAATAGATTGAGTCTCCTCCAAGTA
ATGCTCTCAGATCTAATAGTTATCCTTCAACCAATCTCAACTTTTCAAACAACCAACA
CATTGATGAATTGGGGTGGTGGTAAATCCAGTGATCAATACCGTGCAGCTTACGGCA
ACATTAACACCCATCAGCAATCACCTTACCACAAAATCATTCAAACCCGAACCGCCGGT
CATCTTTGATCCGTTTGAAGATCATCTTCATCACATTCTCCACAAATAAATCTTGATC
ATATCGGAATCAAGAACATCATCAGTCACCAAGTGCCATCTTTACCGGCTGAAACAATCA
AGGCAATCACGACAGATCCAAGTTCCAATCGGCTTTGGCGACAGCTCTATCTTCCATCA
TGGGCGGCGGATTTAAAGATTGATCACAATGTGACTAGAAATGAAGCTGAGAAGAGCCCTT

AAAGAGAATTGTTATATATATGTTCTTATATACTCAGTACATTGGTAAATGGGTTTAGAC
TTTCACTAGTTTCTTAGTTTCATCTATATATTGGTTGTTAATCACAAGTTTATTTTGTG
TTGGAGTTTATGGAACTAATGTGTACATATGAACTTTAGAACGAATAAATAAACTTGG
AATTCCTTTTTAAAAA

>G2110 Amino Acid Sequence (conserved domain in AA coordinates:239-298)

MEKDDFLRSGHGREESHDEMRKLDSSHDDSHQEHDIIRSKLDSTKVEMDEAKEENRRLK
SSLSKIKKDFDILQTQYNQLMAKHNEPTKFQSKGHHQDKGEDEDREKVNREELVSLSLG
RRLNSEVPSGSNKEEKNKDVÉEAEGRNYDDNEKSSIQGLSMGIEYKALSNPNEKLEIDH
NQETMSLEISNNNKIRSONSFQFKNDGDDHEDEDEILPQNLVKKTRVSVRSRCETPTMND
GCQWRKYGQKIAKGNPCPRAYRCTIAASCPVRKQVQRCESEDMSILISTYEGTHNHPLPM
SATAMASATSAASMLLSGASSSSAAADLHGLNFSLSGNNTPKPKTHFLQSPSSSGHP
TVTLDLTSSSSQPFLLSMLNRFSSPPSNVSRNSYPSNLNFSNNTNTLMNWGGGNPS
DQYRAAYGNINTHQQSPYHKIIQTRTAGSSFPDFGRSSSSHSPQINLDHIGIKNIISHQV
PSLPAETIKAITTDPFQALATALSSIMGGDLKIDHNVTRNEAEKSP*

>G2442 (71..997)

TCGACCAATTTAGACCAATCCAAATTCGTCGTCCTTTCTCTGTGTAGTCTAATTATATA
TTACAAGTAGATGAATTGGTTACCTGAAGCTGAAGCTGAGGAGCACTTGAAAGGTATTCT
CTCTGGTGATTCTTTGATGGTCTCACCAATCACCTTGATTGCCCACTTGAAAGACATCGA
TTCCACCAATGGTGAGGAGATTGGGTGCGCAGGTTTCAAGACCTTGAGCCTCCTCCCTT
GGATATGTTCCCTGCTTTGCCCTCTGACCTCACCTCTTGTCCTCAAGGCGCCGCTCGTGT
GCGGATTCCCAACAACATGATTCTGCTTTGAAGCAGTCCTGTTCTTCTGAAGCCTTGTC
CGGCATTAATAGCACTCCCCACCAATCTTCAGCTCCTCTGATATCAAAGTTTCATATCT
ATTTCACTCTCTAATCCAGTGTCTGAGTTCGAGAACAGTTATGGTTCTCTCTCCACCCA
AAACTCCGGATCTCAGAGATTGGCTTTCCCTGTGAAAGGCATGAGAAGCAAGCGCAGACG
CCCCACAACAGTGAGACTTAGCTACCTTTTCCCTTTGAACCCAGAAAGTCAACTCCGGG
TGAATCAGTAACCGAGGGTTACTATTCTTCTGAGCAACATGCCAAGAAGAAGCGCAAGAT
TCATCTGATCACCCACACCGAGTCTTCCACTTTGGAGTCAAGTAAGTCGGATGGGATAGT
CCGGATATGCACATCTATTGTGAGACAATCACGACCCACAGTGGAGGCAAGGACCCAGTGG
ACCCAAGACCTCTGCAACGCTTGCGGAGTCCGGTTCAAATCTGGTCGCTAGTTCCAGA
ATACCGGCCAGCCTCAAGCCCGACCTTCATCCCATCTGTGCATTCAAACCTCACACAGGAA
GATCATTGAGATGAGAAAGAAGGACGACGAGTTTGATAACAGCATGATTGCGAGTGATAT
CCAGAAGGTAAAGCAGGGGAGGAAGAAATGGTATAAAAGTA

>G2442 Amino Acid Sequence (domain in aa coordinates: 220-246)

MNWLPEAAEAEHLKGILSGDFFDGLTNHLDCPLEDIDSTNGEGDWVARFQDLEPPPLDMF
PALPSDLTSCPKGAARVRIPNNMIPALKQSCSSEALSGINSTPHQSSAPPDIKVSYLEFQS
LTPVSVLENSYGLSTQNSGSQRLAFPVKGMRSKRRRPTTVRLSYLFPFEPKSTPGESV
TEGYYSSEQHAKKRIHLITHTESSTLESSKSDGIVRICTHCETITTPQWRQGPSGPKT
LCNACGVRFKSGRLVPEYRPASSPTFIPSVHSNSHRKIIEMRKKDDEFDTSMIRSDIQKV
KQGRKKMV*

>G1051 (66..1031)

CCTGTAAATTCAGATTTGCTTTCTTTGGTAATCTTTTGGATCAAGATCCATCTATTTT
CTTCAATGGCACAACCTCCCTCCTAAATCCCCAACATGACACAACATTGGCCTGATTTCT
CTTCCCAAAAGCTCTCTCCTTTCTTACCCCAACCGCAACCGCTGTGCGCACCGCTACAA
CCACCGTACAAAACCCCTCATGGGTGACGAATTCTCTGACTTCTCAGCGTCTCGCCGTG
GCAACCACCGTCGTTCCATCAGCGACTCTATCGCATTCTCGAAGCTCCAACAGTCAGCA
TCGAAGACCACCAATTGACAGGTTGATGACGAACAGTTCATGTCGATGTTACCGGACG
ACGACAACCTTCATAGCAATCCTTCCCATATCAACAACAAAAATAACAATGTGGGGCCCA
CGGGATCTTCTCTGAACACATCCACGCCGTCCAATAGCTTCAACGACGATAACAAAGAAT
TACCACCGTCCGATCATAACATGAACAATAATATCAACAACAATAACGATGAAGTCC
AAAGCCAATGCAAGATGGAGCCAGAAGATGGTACGGCGTCAATAACAATTCGGGTGATA
GCTCCGGCAACCGGATTCTCGATCCCAAAAGGGTTAAGAGAATATTAGCAAAATCGGCAAT
CAGCACAGAGATCAAGGGTGAGGAACTGCAATACATACAGAGCTCGAACGTAGCGTCA
CTTCGTTGACGGCGGAAGTGTCTAGTGTATCGCCAAGAGTTGCATTCTTGGATCATCAAC
GTTTGCTTCTTAACGTTGACAACAGCGCTCTCAAGCAACGAATCGCTGCTTTATCTCAAG
ACAAGCTTTTCAAAGACGCACATCAAGAAGCATTGAAGAGAGAAATAGAGAGACTTCGAC
AAGTGTATAATCAACAAGCCTCACGAATGTGGAAAATGCAATCATTTATCGGCGACCG
GAGCCGGTGCTACTCCGGCCGTGACATCAAGTCGTCGGTTGAAACAGAGCAGCTCCTCA

ATGTCTCATAAATTAACCATCATGCATCATCAACATTTCTCTCTTTAGCTTCTTGG
CAAAAGTTCTTGACTATAAAATCTCTTTTCGGGTAAGAAATTCAGGAGATATACATTTTTT
ATTCTAATCACATTGTTTTTAAGTTGTGATGAATTCAGTTTGATGTATCTTATTTATTTT
GTTTATGTCGCTCTTTTTTTCTTGGGGTTGATGGAAGGAATCATCAATTGTTGTTGTAC
AAAGAACTAGTTGAATTTTTTTTTTTTTTTT

>G1051 Amino Acid Sequence (domain in AA coordinates 189-250)
MAQLPPKIPNMTQHWPDFSSQKLSPFSTPTATAVATATTTVQNPSWVDFLDFSASRRGN
HRRSISDSIAFLEAPTVSIEDHQFDRFDDEQFMSMFTDDDLHLSNPISHINNKNVVGPTG
SSSNTSTPSNSFNDDNKELPPSDHNMNNINNNYNDEVQSQCKMEPEDGTASNNNSGDSS
GNRILDPKRVKRILANRQSAQRSVRKQLQYISELERSVTSLQAEVSVLSPRVAFLDHQRL
LLNVDNSALKQRIALSQDKLFKDAHQEALKREIERLRQVYNQQLTNVENANHLSATGA
GATPAVDIKSSVETEQLLNVS*

>G1052 (138..1127)
TGATCATCTAAAACCTTCAATTTCTCTCTTGATCCTCAGTTGAATTTTTTGTGTTTCTC
TCAAATCTTTGATCCTTTCTCTTTGTTTTTCATTTGACCTCTTACAAAAAATCTGGTGTG
CCATTAAATCTTTATTAATGGCACAACCTTCTCCGAAAATCCCAACCATGACGACGCCAA
ATTGGCCTGACTTCTCTCCAGAACTCCCTTCCATAGCCGCAACGGCGGCAGCCGCGAG
CAACCGCTGGACCTCAACAACAAAACCTTCATGGATGGATGAGTTTCTCGACTTCTCAG
CGACTCGCCGTGGGACTCCCGTCTTATAAGCGACTCCATTGCTTTCTTGAACCACT
CTTCTCCGGCGTCCGAAACCACCACTTCGATAGGTTTGACGACGAGCAATTCATGTCCA
TGTTCAACGACGAGCTACACAACAATAACCACAATCATCATCATCACAGCATCAACG
GCAATGTGGGTCCACGCGTTTATCTCCAACACCTCCACGCGTCCGATCATAATAGCC
TTAGCGACGACACAACAACAAGAAGCACCACCGTCCGATCATGATCATCACATGGACA
ATAATGTAGCCAATCAAAACAACGCCGCGGTAACAATTACAACGAATCAGACGAGGTCC
AAAGCCAGTGCAAGACGAGCCACAAGATGGTCCGTCGCGCAATCAAAACTCCGGTGGAA
GCTCCGGTAATCGTATTCACGACCTAAAAGGGTAAAAGAATTTTAGCAAATAGGCAAT
CAGACAGAGATCAAGGGTGAGGAAATGCAATACATATCAGAGCTTGAAAGGAGCGTTA
CTTCATTGCAGACTGAAGTGTCTAGTGTATCGCCAAGAGTTGCGTTTTTGGATCATCAGC
GATTGCTTCTCAACGTCGACAATAGTGCTATCAAGCAACGAATCGCAGCTTTAGCACAAG
ATAAGATTTTCAAAGACGCTCATCAAGAAGCATTGAAGAGAGAAATAGAGAGACTTCGAC
AAGTATATCATCAACAAAGCCTCAAGAAGATGGAGAATAATGTCTCCGATCAATCTCCGG
CCGATATCAAAACCGTCCGTTGAGAAGGAACAGCTCCTCAATGTCTAAAGCTGTTGTTCA
CTAAGATCTTCTTTTTCATGGCGAAAAGATTCTTGACTATAAAACCTCTTTGTGTCAAGA
AATTAATTTATCAAAGAAGATGGCCTTTTTTATTGTATCAATCACATTTTTTTAAGTTG
TGATGAATTTGCTTTTGATGTATCTGTTTTTTTTTTTTTTTTT

>G1052 Amino Acid Sequence (domain in AA coordinates 201-261)
MAQLPPKIPTMTTPNWPDPSSQKLSIAATAAAAATAGPQQNPSWMDFLDFSATRRGT
HRRSISDSIAFLEPPSSGVGNHFDHFDDEQFMSMFDNDVHNNNNHHHHHSINGNVGPT
RSSSNTSTPSDHNLSDDDNKEAPPDHDHMDNNVANQNNAAAGNNYNESDEVQSQCKT
EPQDGPANQNSGSSGNRIHDPKRVKRILANRQSAQRSVRKQLQYISELERSVTSLQTE
VSVLSPRVAFLDHQRLLLNVDNSAIKQRIALAQDKIFKDAHQEALKREIERLRQVYHQQ
SLKKMENNVSQSPADIKPSVEKEQLLNVS*

>G1079 (1..1995)
ATGGGTTGTGCTGCTTCAAGAATTGATAATGAAGAAAAGGTTTTAGTGTGTAGGCAGAGA
AAGAGGCTAATGAAAAGTTATTAGGGTTTCAGGGGAGAATTTGCAGATGCACAGTTGGCT
TATCTTAGAGCTTTGAGGAACACTGGTGTTACTCTTAGGCAATTCAGTGTCTGAGACC
TTGGAGCTTGAAAACACTAGTTATGGTTTAAGTTTGCCTTTGCCCTCCTTCGCCTCCTCCT
ACATTGCCTCCTTCACCTCCACCACCTCCTCCATTAGCCCGGATTTGAGAAATCCTGAG
ACTAGTCATGACTTGGCTGATGAGGAGGAAGAGGGTGAAAATGATGGTGGTAATGATGGA
AGTGGTGACGCTCCTCCGCTCCATTGCCGAATTTCTTGAACATTTGGAACCTTTTGAG
TCCTTGAGCTGCATAGTCAATCCAAATGGTGACAATGTAGTTACACAAGTTGAAGTGAAG
AAGAAACAACAATTCAGCAAGCTGAAGAGGAAGATTTGGGCGGAGACGAAGTCTCAATTT
GAGGAAGAAGATGAGCAACAAGAAGCAGGAGGTACTTGCCTTGATTTGAGTGTTCATCAA
ATAGAGGCTGTTAGTGGCTGTAAACATGAAGAAGCCACGTCGCTGAAAGTTAAGCTGGGA
GAAGTTATGACGGTAACCTCATCTATGACAAGCTGCTCCGGTAAAGATCTTGAGAAAAC
CATGTGACTGATTGTAGAATCAGGAGGACCTTAGAAGGAATCATCAGAGAGTTGGATGAT
TATTTCTTAAAGCATCGGGTTGCGAGAAGGAGATAGCTGTGATAGTAGACATCAACAGT

AGGGATACTGTTGATCCTTTTCAGGTACCAGGAAACAAGAAGGAAGAGAAGCAGCTCGGCA
AAGGTATTTCAGTGCATTGTTCATGGAGTTGGTCTTCAAAGTCTCTTCAGTTGGGCAAAGAT
GCTACAACAAGCGGGACTGTTGAACCTGTAGGCCTGGAGCTCACTGCAGCACACTTGAG
AAGCTATACACAGCTGAGAAGAACTTTACCAGCTAGTCAGAAACAAGAGATTGCCAAA
GTGGAGCATGAGAGGAAGTCTGCATTACTGCAAAAGCAAGATGGGGAAACCTATGATTTG
AGCAAAATGGAGAAAGCACGCTTGTCTTTGGAGAGTTTGGAAACCGAGATACAGCGTCTA
GAAGATTCCATAACTACAACACGCTCATGTTTGCTTAACCTTGATCAATGATGAGCTGTAT
CCGAGCTAGTTGCTTTAACTTCAGGGCTAGCACAGATGTGGAAAACAATGCTCAAGTGT
CATCAAGTTCAAATTCATATATCCCAGCAACTGAACCATCTTCCGGATTACCCGAGTATA
GATCTCAGTTCCGAATACAACGCCAGGCGGTTAATGAACTAGAGACCGAGGTTACTTGC
TGGTACAATAGCTTTTGCAAGTTAGTAAATTCAGCGAGAATACGTGAAAACACTCTGT
ACGTGGATCCAACCTTACTGATCGCCTCTCTAACGAAGACAACCAAGAAGTAGCTTGCCT
GTTGCTGCTCGTAAGCTCTGCAAGAGTGGCAGCTTGAATACAACCTGCGTAGGAAATGC
AATAAACTTGAGAGGAGGCTTGAGAAAGAGCTAATTTCACTGGCTGAGATTGAAAGAAGG
CTCGAGGGGATTTTAGCAATGGAAGAGGAGGAAGTAAGCTCAACGAGTTTGGGCTCTAAG
CATCCGTTGTCAATCAAACAAGCCAAGATCGAAGCCTTGAGAAAACGAGTGGATATTGAG
AAAACCTAAGTACTTAACTCGGTCCGAGTTAGTAAGAGAATGACACTAGACAACCTCAAA
TCAAGCCTTCCCAATGTCTTTAGATGTTGACTGCTCTAGCTAATGTCTTTGCCAATGGG
TTTGAATCCGTTAATGGCCAAACCGGTACAGATGTTCCGACACATCCCAACATTCCGAT
GAATCTCAACCTAA

>G1079 Amino Acid Sequence (conserved domain in AA coordinates:1-50)

MGCAASRIDNEEKVLVCRQRKRLMKLLGFRGEFADAQLAYLRALRNTGVTLRQFTESET
LELENTSYGLSLPLPPSPPTLPPSPPPPPPPFSPDLRNPETSHDLADEEEEGENDGGNDG
SGAAPPPLPNSWNIWNPESLELHSHPNGDNVVTQVELKKKQIQQAEEEDWAETKSQF
EEEDEQQEAGGTCLDLSVHQIEAVSGCNMMPRLKFKLGEVMDGNSMSTSCSGKDLEKT
HVTDCRIRRTLEGIIRELDDYFLKASGCEKEIAVIVDINSRDTVPDPFRYQETRRKRSSSA
KVFSALSWSWSSKSLQLGKDATTSGTVEPCRPGAHCSTLEKLYTAEKKLYQLVRNKEIAK
VEHERKSALLQKQDGETYDLSKMEKARLSLESLETEIQRLEDSITTTTRSCLLNLINDELY
PQLVALTSGLAQMWKTMKCHQVQIHISQQLNHLDPYPSIDLSSEYKQAVNELETEVTC
WYNSFCKLVNSQREYVKTLCTWIQLTDRLSNEDNQRSSLPVAARKLCKEWQLEYNLRRKC
NKLERRLEKELISLAEIERRLEGILAMEEEVSSLSGSKHPLSIKQAKIEALRKRVDIE
KTKYLSNEVSVKRMTLNKLSSLPNVFQMLTALANVFANGFESVNGQTGTDVSDTSQHS
ESQP*

>G1335 (56..667)

TTTTTTTTTAAAGATTAGAGAGAAAAGTGAGTTATTAAGAGATTCCAATCAAAATGAG
CGGAGACAACGGCGGTGGTGAGAGGCGCAAAGGCTCCGTCAAGTGGTTTGATACCCAGAA
GGGTTTCGGCTTCATCACTCCTGACGACGGTGGCGACGATCTCTTCGTTACCAAGTCTC
CATCAGATCTGAGGGTTTCCGTAGCCTCGCTGCCGAAGAAGCCGTAGAGTTCGAGGTTGA
GATCGACAACAACACCGTCCCAAGCCATCGATGTTTCTGGACCCGACGGCGCTCCCGT
CCAAGGAAACAGCGGTGGTGGTTCATCTGGCGGACGCGCGGTTTCCGTGGAGGAAGAGG
AGGTGGACGCGGATCTGGAGGTGGATACGGCGGTGGCGGTGGTGGATACGGAGGAAGAGG
AGGTGGTGGTTCGAGGAGGACGCACTGCTACAAGTGTGGTGGAGCCCGGTACATGGCGAG
AGACTGTTCTGAAGGCGGTGGAGGTTACGGAGGAGGCGCGGTGGCTACGGAGGTGGAGG
CGGATACGGCGGAGGAGGTGGTGGTTACGGAGGTGGTGGCGGTGGAGGTGGTGGCGCGG
GGGAAGCTGCTACAGCTGTGGCGAGTCGGGACATTTCCGACGGGATTGCACCAAGCGGTGG
ACGTTAAAACCAACGCCGGTTACGCGGTGGAGAAGAGTGAGTTGGTTATCTCACAAGTGA
TCGGTTCTTTCTCCGCGCGCTTCTATCTCTTATTATCCACTTTTGTCTTATTATGATG
GATCTCTATCTTTGTTAGTTGGTTTTTCTTGATGGTTTCGGATTAGGACTCTCTTTTG
GTTTTGCTACTTATGGTTGGTTTTATTTATGGTACTTGTGATATGGGTGAAATGCTCTAC
TTGTTGCTCTGTTTCAAGTGTTCATAATATGCGAACAATATTCTGGGTTTTGTTTCAAA
AAAAA

>G1335 Amino Acid Sequence (domain in AA coordinates: 24-43, 131-144, 185-203)

MSGDNGGERRKGSVKWFDQKGFGITPDDGGDDLFDVHQSSIRSEGFRSLAAEEAVEFE
VEIDNNNRPKAIDVSGPDGAPVQNSGGGSSGGRGFGGGRGGGRGSGGGYGGGGGGYGG
RGGGGRGSGDCYKCGEPGHMARDCEGGGGYGGGGGGYGGGGYGGGGGGYGGGGGGGG
GGGSCYSCGESGHFARDCTSGGR*

>G157 (31..621)

GGGCATAACCCCTTATCGGAGATTTGAAGCCATGGGAAGAAGAAAAATCGAGATCAAGCGA
ATCGAGAACAAAAGCAGTCGACAAGTCACTTTCTCCAAACGACGCAATGGTCTCATCGAC
AAAGCTCGACAACCTTTCGATTCTCTGTGAATCCTCCGTCGCTGTTGTCTCGTATCTGCC
TCCGGAACCTCTATGACTCTTCTCCGGTGACGACATTTCCAAGATCATGATCGTTAT
GAAATACAACATGCTGATGAACCTTAGAGCCTTAGATCTTGAAGAAAAAATTCAGAATTAT
CTTCCACACAAGGAGTTACTAGAAAAGTCCAAAGCAAGCTTGAAGAACCAATGTTCGAT
AATGTAAGTGTAGATTCTCTAATTTCTCTGGAGGAACAACTTGAGACTGCTCTGTCCGTA
AGTAGAGCTAGGAAGGCAGAACTGATGATGGAGTATATCGAGTCCCTTAAAGAAAAGGAG
AAATTGCTGAGAGAAGAGAACCAGGTTCTGGCTAGCCAGATGGGAAAGAATACGTTGCTG
GCAACAGATGATGAGAGAGGAATGTTTCCGGGAAGTAGCTCCGGCAACAAAATACCGGAG
ACTCTCCCGCTGCTCAATTAGCCACCATCATCAACGGCTGAGTTTTCACCTTAAACTCAA
AGCCTGATTACATAAATTAAAGAGAATAAATTTGTATATTATAAAAAGCTGTGTAATCTCAA
CCTTTTATCTTCTCTAGTGTGGAATTTAAGGTCAAAAAGAAAACGAGAAAGTATGGATC
AGTGTGTACCTCCTTCGGAGACAAGATCAGAGTTTGTGTGTTTGTGTCTGAATGTACGG
ATTGGATTTTTAAAGTTGTGCTTTCTTTCTTCAAAAAAAAAA
>G157 Amino Acid Sequence (domain in AA coordinates: 2-57)
MGRRKIEIKRIENKSSRQVTFKRRNGLIDKARQLSILCESSVAVVVVSASGKLYDSSSG
DDISKIIDRYEIQHADELRLDLEEKIQNYLPHKELLETVQSKLEPNVDNVSVDSLISL
EEQLETALSVSRARKAELMMEYIESLKEKEKLLREENQVLASQMGKNTLLATDDERGMFP
GSSSGNKIPETLPLLN*
>G1895 (1..954)
ATGAATAACCAATCTGTTACTGACAATACAAGTCTTAAGCTGTCTATCTAATCTTAACAAC
GAGTCAAAAGAAACATCTGAGAACAGTGATGACCAACACAGCGAGATCACAACAATTACA
TCGGAAGAAGAGAAAACAACTGAAGTGAAGAAACAGACAAGATTCTTCCATGTCCGAGA
TGCAACAGCGCAGACACCAAACTTCTGTTACTACAACAACCTACAACGTTAACCAGCCACGT
CACTTCTGTAGAAAATGCCAGAGGTATTTGGACCGCTGGTGGATCCATGAGGATCGTCCCG
GTTGGCTCAGGCCGTCGCAAGAACAAGGGATGGGTTTCTTCAGACCAGTACCTGCACATC
ACTTCCGAGGATACTGACAATTACAATAGCTCCTCAACAAGATTCTAAGCTTCGAGTCT
TCGGACTCTTTGGTAACTGAGAGGCCCTAAGCATCAATCAACGAAGTGAAGATAAACGCT
GAACCTGTTTCACAAGAACCAACAACCTTCCAAGGGTTACTTCTCCCAAGCATCCCT
GTTTCGCCTCCTTGGCCTTACCAATACCCTCCAAACCCTAGTTTCTACCACATGCCCGTC
TACTGGGGCTGCGCGATACCGGTTTGGTCTACCCTCGACACTTCTACATGTCTTGGGAAA
AGGACAAGAGACGAAACTTCTCATGAACTGTTAAAGAGAGTAAAAATGCTTTTGAGAGA
ACAAGCTTGCTTTTGAATCTCAGAGCATCAAAAATGAAACAAGTATGGCTACAAATAAC
CATGTGTGGTATCCAGTACCGATGACCCGCGAGAAGACACAAGAATTCAGCTTTTTCAGT
AATGGAGCTGAAACAAGAGCAGCAACAACAGATTGTCCTCGAAACGTATCTTAACCTG
CAAGCAAACCTGCAGCCATGGCAAGATCTATGAACCTCAGAGAGAGCATATAA
>G1895 Amino Acid Sequence (domain in AA coordinates: 55-110)
MNNQSVTDNTSLKSSNLNNESENKSENDDQHSEITTTTSEBEKTELKPKDKILPCPR
CNSADTKFCYNNYNVNQPRHFCRKQRYWTAGGSMRIVPVGSGRRKNKGWVSSDQYLHI
TSEDTDNYSSTKILSFESSDSLVTFRPKHQSNEVKINAEPVSQEPNNFQGLLPPQASP
VSPWPYQYPPNPSFYHMPVYWGCAIPVWSTLDTSTCLGKRTRDETSHETVKESKNAFER
TSLLESQSIKNETSMATNNHVWYPVPMTREKTQEFSSFFSNGAETKSSNNRFVPETYLN
QANPAAMARSMNFRSI*
>G1900 (1..897)
ATGCTGGAACTAAAGATCCTGCGATAAAGCTCTTTGGTATGAAAATTCCTTTCCCGACG
GTTTTAGAGGTTGCTGATGAAGAAGAAGAAAAGAACCAAAACAAGACATTAAGTATCAA
TCGGAGAAAGACAAACCCCTAAAGAAACCAACCAAGATTCTTCCATGTCCAAGATGCAAC
AGCATGGAGACTAAGTTCTGTTACTACAACAACCTACAACGTAAACCAACCTCGCCATTTT
TGTAAGCTTGTGAGAGATATTGGACCTCAGGTGGGACCATGAGAAGTGTTCGAATCGGA
GAGGACGGCGCAAGAACAAGAACTCACCACCTTACATTACCACCATGTGACTATC
TCCGAAACAATAGTCCGGTCTTAGTTTTCAGCCTCGGAGATGATCAAAAGGTCTCGAGT
AATAGGTTTGGTAACTCAAAAGCTAGTTTTCAGGATAGAGAACAATGACGAGCGCTCTAAT
AACAACTTTCGAACGGTTTGAATTGTTTCCGGGAGTTTCGTGGCCGTACACGTGGAAT
CCTGCGTTTACCCGTTTACCCTTATTGGAGCATGCCAGTGTGTCTTCTCCGGTAAGT
TCAAGTCCCTACTTCTACTCTTGGTAAGCATTGAGAGACGAAGACGAGACGGTGAAGCAA
AAACAGAGGAATGGATCTGTATTGGTTCCAAAGACTTTGAGAATTGATGATCCTAATGAA

GCTGCAAAGAGTTTCGATATGGACAACACTTGGGATCAAGAACGAAGTTATGTTCAATGGG
TTTGGTTGCAAGAAAGAGGTTAAGCTCAGTAACAAAGAAGAACAGAGACCTCACTTGT
CTTTGTGCAAACCGTGCCTTATCAAGATCAATCAATTTCCATGAGCAGATGTGA
>G1900 Amino Acid Sequence (domain in AA coordinates: 54-106)
MLETKDPAIKLFGMKIPFPTVLEVADEEEKNQKLTLDQSEKDKTLKKPTKILPCPRCN
SMETKFCYNNYNNVNPQRHFKACQRYWTSGGTMRSPVIGARRKNKNSPTSHYHHVTI
SETNGPVLFSFSLGDDQKVSSNRFNGNQLVARIENNDERSNNNTSNGLNCFPVSWPYTWN
PAFYVPVYPWMPVLSSPVSSSPTSTLKGHSRDEDETVKQKQRNGSVLVPKTLRIDDPNE
AAKSSIWTTLGIKNEVMFNGFGSKKEVKLSNKEETETSLVLCANPAALSRSINFHEQM*
>G2007 (1..861)
ATGGGAAGGCAGCCATGTTGTGACAAGCTCATGGTGAAGAAGGGGCCGTGGACGGCGGAG
GAAGACAAGAAACTGATAAACTTTATCTTGACCAACGGCCACTGTTGCTGGAGGGCTTTG
CCGAAGCTGGCCGGTCTCCGCTCGCTGTGGGAAGAGCTGCCGTCTACGGTGGACCAATTAT
CTCCGACCTGACTTGAAGAGAGGTCTTCTCTCCGACGCCGAGGAACAGCTTGTCAATCGAC
CTTCATGCTCTTCTCGGCAACAGATGGTCCAAGATCGCTGCAAGATTACCAGGAAGAACA
GACAACGAAATAAAAAATCATTGGAATACTCATATCAAGAAGAAGCTCCTTAAGATGGAA
ATCGATCCTTCGACCCATCAACCTTTTAAACAAAGTATTTACCGATACAACTTAGTCGAT
AAATCTGAAACTTCATCGAAAGCCGACAATGTAAATGATAATAAAATCGTAGAGATCGAT
GGGACAACGACAAATACATAGATGATAGCATTATCACTCATCAAAATAGTTCAAATGAT
GATTATGAATTACTTGGTGATATAATTCAATAATTATGGAGATTTATTTAATATTCTATGG
ACCAACGATGAACCTCCTCTAGTCGATGATGCATCATGGAGCAATCATAACGTTGGTATT
GGAGGAACAGCTGCAGTTGCAGCCTCAGACAAGAACAACACTGCTGCCGAGGAAGATTTC
CCGGAAGATCATTGAAAAACAGAACGGCGAAAGTTGGATGTTCTTGGATTATTGCCAA
GAATTTGGTGTGAAGATTTTGGGTTTCGAGTGTTACCATGGTTTGGTCAAAGCTCCATG
AAGACGGGTCAACAAGGACTAG
>G2007 Amino Acid Sequence (domain in AA coordinates: TBD)
MGRQPCCDKLMVKGPWTAEEEDKKLINFILTNHCCWRALPKLAGLRRCGKSCRLRWNTY
LRPDLKRGLLSDAEEQLVIDLHALLGNRWSKIAARLPGRDNEIKNHNWTHIKKLLKME
IDPSTHQPLNKVFTDNLVLDKSETSSKADNVNDNKIVEIDGTTNTIDDSIITHQNSSND
DYELLGDIHNYGDLFNILWTNDEPPLVDDASWSNHNVGIGTAAVAASDKNNTAAEEDF
PERSFEKQNGESWMFLDYCQEFVGVEDFGFECYHGFQSSMKTGHKD*
>G214 (238..2064)
TGAGATTTCTCCATTTCCGTAGCTTCTGGTCTCTTTTCTTTGTTTCATTGATCAAAAGCA
AATCACTTCTTCTTCTTCTTCTTCTCGATTTCTTACTGTTTCTTATCCAACGAAATCTG
GAATTAATAAGGAATCTTTATCGAATCCAAGCTGATTTTGTCTTTTCATTGAATCATC
TCTCTAAAGTGGAATTTGTAAAGAGAAGATCTGAAGTTGTGTAGAGGAGCTTAGTGATG
GAGACAAATTCGTCTGGAGAGATCTGGTTATTAAGACTCGGAAGCCATATACGATAACA
AAGCAACGTGAAAGGTGGACTGAGGAAGAACATAATAGATTCAATGAAGCTTTGAGGCTT
TATGGTAGAGCATGGCAGAAGATTGAAGAACATGTAGCAACAAAACTGCTGTCCAGATA
AGAAGTCACGCTCAGAAATTTTCTCCAAGGTAGAGAAAGAGGCTGAAGCTAAAGGTGTA
GCTATGGGTCAAGCGCTAGACATAGCTATTCTCTCCACGGCCTAAGCGTAAACCAAAC
AATCCTTATCTCGAAAGACGGGAAGTGAACGATCCTTATGTCAAAAACGGGTGTGAAT
GATGGAAAAGAGTCCCTTGGATCAGAAAAAGTTCGCATCCTGAGATGGCCAATGAAGAT
CGACAACAATCAAAGCCTGAAGAGAAAACTCTGCAGGAAGACAACCTGTTTCAAGATTGTTT
ACTCATCAGTATCTCTCTGCTGCATCCTCCATGAATAAAAGTTGTATAGAGACATCAAAC
GCAAGCACTTTCCGCGAGTTCTTGCTTACGGAAGAGGGAAGTCAGAATAACAGGGTA
AGAAAGGAGTCAAACCTCAGATTTGAATGCAAAATCTCTGGAACCGGTAATGAGCAAGGA
CCTCAGACTTATCCGATGCATATCCCTGTGCTAGTGCCATTGGGGAGCTCAATAACAAGT
TCTCTATCACATCCTCCTTCAGAGCCAGATAGTCATCCCCACACAGTTGCAGGAGATTAT
CAGTCGTTTCTTAATCATATAATGTCAACCTTTTACAAACACCGGCTCTTTTACTTGCC
GCAACTTTTCGCTCATCATTTTGGCCTCCCGATTCTAGTGGTGGCTCACCTGTTCCAGGG
AACTCACCTCGAATCTGGCTGCCATGGCCGAGCCACTGTTGCAGCTGCTAGTGCTTGG
TGGGCTGCCAATGGATTATTACCTTTATGTGCTCCTCTTAGTTCAAGGTGGTTTCACTAGT
CATCCTCCATCTACTTTTGGACCATCATGTGATGTAGAGTACACAAAAGCAAGCACTTA
CAACATGGTTCTGTGCAGAGCCGAGAGCAAGAACAACCTCCGAGGCATCAAAGGCTCGATCT
TCACTGGACTCAGAGGATGTTGAAAATAAGAGTAAACCAGTTTGTGATGAGCAGCCTTCT
GCAACACCTGAGAGTGATGCAAAGGTTTCAGATGGAGCAGGAGACAGAAAACAAGTTGAC

CGGTCCTCGTGTGGCTCAAACACTCCGTCGAGTAGTGATGATGTTGAGGCGGATGCATCA
GAAAGGCAAGAGGATGGCACCACCAATGGTGAGGTGAAAGAAACGAATGAAGACACTAATAAA
CCTCAAACCTTCAGAGTCCAATGCACGCCAGTAGAATCAGCTCCAATATAACCGATCCA
TGGAAGTCTGTGTCTGACGAGGGTCTGAATTGCCTTCCAAGCTCTCTTCTCCAGAGAGGTA
TTGCCGCAAAGTTTTACATATCGAGAAGAACACAGAGAGGAAGAACAACAACAAGAA
CAAAGATATCCAATGGCACTTGATCTTAACCTCACAGCTCAGTTAACACCAGTTGATGAT
CAAGAGGAGAAGAGAAACACAGGATTTCTTGGAATCGGATTAGATGCTTCAAAGCTAATG
AGTAGAGGAAGAACAGGTTTTAAACCATACAAAAGATGTTCCATGGAAGCCAAAGAAAGT
AGAATCCTCAACAACAATCCTATCATTCATGTGGAACAGAAAGATCCCAAACGGATGCGG
TTGGAAACTCAAGCTTCCACATGAGACTCTATTTTCATCTGATCTGTTGTTTGTACTCTG
TTTTTAAGTTTTCAAGACCACCTGCTACATTTCTTTTCTTTTGAGGCCTTTGTATTTGT
TTCTTGTCCATAGTCTTCTGTAACTTTGACTCTGTATTATTCAACAAATCATAAACT
GTTTAATCTTTTTTTTTTCCA

>G214 Amino Acid Sequence (domain in AA coordinates: 22-71)
METNSSGEDLVIKTRKPYTITKQRRWTEEEHNRFI EALRLYGRAWQKIEEHVATKTAVQ
IRSHAQKFFSKVEKEAEAKGVAMGQALDIAIPPPRPKRKPNPYPRKTGSGTILMSKTGV
NDGKESLGSEKVSHPMANEDRQSKPEEKLQEDNCSDFTHQYLSAASSMNKSCIETS
NASTFREFLPSREEGSQNNRVRKESNDLNAKSLENGNEQGPQTYPMHIPVLVPLGSSIT
SSLSHPPSEPDSPHTVAGDYQSFNHHMSTLLQTPALYTAATFASSFPPDSSGGSPVP
GNSPPNLAAMAAATVAAAASAWWAANGLPLCAPLSSGGFTSHPPSTFGPSCDVEYTKAST
LQHGVSQSREQEHSEASKARSSLDSEDEVENKSKPVCHEQPSATPESDAKSGDAGDRKQV
DRSSCGSNTPSSDDVEADASERQEDGTNGEVKETNEDTNKPQTSSESNARRSRISNITD
PWKSVSDEGRIAFQALFSREVL PQSFYREEHREEEQQQEQRYPMALDLNFTAQLTPVD
DQEEKRNTGFLGIGLDASKLMSRGRGTGFKPKRCSMEAKESRILNNNP I IHVEQKDPKRM
RLETQAST*

>G2155 (63..740)

CTCATATATACCAACCAACCTCTCTCTGCATCTTTTATTAACACAAAATTCCAAAAGATT
AAATGTTGTGCGAAGCTCCCTACACAGCGACACTTGACCTCTCTCCCTCCTCTCCCTCCA
TGGAACCGTCCGGGCGTCCACGTGCGACACTCGAGGTTCCAAAACAAACCTAAAGCTC
CAATCTTTGTGACCATTGACCCCTCTATGAGTCCTTACATCCTCGAAGTGCCATCCGGAA
ACGATGTGCTTGAAGCCCTAAACCGTTTCTGCCGCGGTAAAGCCATCGGCTTTTGCGTCC
TCAGTGGCTCAGGCTCCGTTGCTGATGTCACCTTTCGCTCAGCCTTCTCCGGCAGCTCCTG
GCTCAACCATTACTTTCCACGGAAAGTTGATCTTCTCTGCTCTCGCCACTTTCTCTCC
CTCCTCTACCTCCTACCTCCTTGTCCCTCCCGTCTCCAATTTCTTCACCGTCTCTCTCG
CCGACCTCAGGGGAAAGTCATCGGTGGATTGCTCGCTGGTCTCTCGTTGCCGCCGGAA
CTGTTTACTTTCGTCGCCACTAGTTTCAAGAACCCTTCTATCACCGGTTACCTGCTACGG
AGGAAGAGCAAAGAACTCGCGCGGAAGGGGAAGAGGAGGGAACATCGCCGCCGGTCTCTG
GAGGTGGTGGAGAGTCGATGTACGTGGGTGGCTCTGATGTCATTTGGGATCCCAACGCCA
AAGCTCCATCGCCGTACTGACCACAAATCCATCTCGTTCAAACCTAGGGTTTCTTCTCTT
TAGATCATCAAGAATCAACAAAAGATTGCATTTTGTAGATTCTTTGTAATATCATAATTG
ACTCACTCTTTAATCTCTCTATCACTTCTTCTTTAGCTTTTCTGCACTGTCAAACCTCA
CATATTTGTAGTTTGATTGACTATCCCCAAGTTTGTATTTTATCATACAAATTTTGTG
CTGCTCTAATGGTTGTTTTTCTGTTGTATAATCTTATGCATTGTTTATTGGAGCTCCA
GAGATTGAATGTATAATATAATGGTTTAAT

>G2155 Amino Acid Sequence (domain in AA coordinates:18-38)
MLSKLPTQRHLHLSPPSPMETVGRPRGRPRGSKNPKAPIFVTIDPPMSPYILEVPSGN
DVVEALNRFCRGAIGFCVLSGSGSVADVTLRQPSPAAPGSTITFHGKFDLLSVSATFLP
PLPPTSLSPVSNFFTVSLAGPQGVIGGFVAGPLVAAGTVYFVATSFKNPSYHRLPATE
EEQRNSAEEGEEGQSPPVSGGGGSEMYVGGSDVIWDPNAKAPSPY*

>G234 (106..1035)

CACAACATCATACCCACCAACATATATAATCTTGATCATAGAGAGATAAACAGAGGCCGC
TATCAAGAACAAGACTAAGAACAAGACTTCACTAGGAGTACAAGTATGGGAAGAGCACCG
TGTTGTGACAAAGCAAACGTGAAGAAAGGCCCTTGCTCTCTGAGGAAGATGCAAACTC
AAATCTTACATTGAAAATAGTGGCACCAGGAGCAATTGGATCGCTTTGCCTCAAAGATT
GGTTTAAAGAGATGTGGAAGAGTTGCAGGCTGAGGTGGCTTAACATCTTAGACCAAAC
ATCAAACATGGTGGCTTCTCTGAGGAAGAAGAAAACATCATTGTAGCCTTTACCTTACA
ATTGGTAGCAGGTGGTCTATAATCGCTGCTCAATTGCCGGGACGAACAGACAACGATATA

AAAACTATTGGAACACGAGGCTCAAGAAGAACTCATTAAACAAACACGCAAGGAGCTT
CAAGAAGCTTGTATGGAGCAGCAAGAGATGATGGTGATGATGAAGAGACAACACCAACAA
CAACAAATCCAAACTTCTTTTATGATGAGACAAGACCAAAATGTTTACATGGCCACTA
CATCATATAATGTTTCAAGTTCAGCTCTTTTCAGAATCAAACCAACTCGTTTTGCGACC
AAGAAGATGTTAAGCCAGTGCTCATCAAGAACATGGTCAAGATCGAAGATCAAGAACTGG
AGAAAACAAACCTCATCATCATCAAGATTCAATGACAACGCTTTTGATCATCTCTCTTTC
TCTCAACTCTTGTTAGATCCTAATCATAACCACCTTAGGATCAGGAGAGGGTTTCTCCATG
AACTCTATCTTGAGCGCCAAACAAACTCTCCATTGCTTAACACAAGTAATGATAATCAG
TGGTTCGGAATTTCCAGGCCGAAACCGTAAACTTGTCTCAGGAGCCTCCACAAGTACT
TCGGCAGATCAAAGCACTATAAGTTGGGAAGACATAAGCTCTCTTGTATTCTGATTCA
AAGCAATTTTTTAAATTATAATAATATATTATCTTAAGATGAAACGTACATCATTATTA
TTAATTGGGGGTACGTAACTATATATGGAATAACGATCTAGTTTGTAAATTTAAAA
>G234 Amino Acid Sequence (domain in AA coordinates: 14-115)
MGRAPCCDKANVKKGPWSPEEDAKLKSYLENSGTGWNIALPQKIGLKRCGKSCRLRWLN
YLRPNIKHGGFSEEEENIICSLYLTIGSRWSIIAQLPGRDNDIKYWNTRLKKLINK
QRKELQEACMEQQEMMVMKQHQQQIQTSMRQDQTMFTWPLHHNVQVPALFRIKP
TRFATKMLSQCSSRTWSRSKIKNWRKQTSSSRFNDNAFDHLSFSQLLLDPNHNHLGSG
EGFSMNSILSANTNSPLNTSNDNQWFGNFQAETVNLFSGASTSTSADQSTISWEDISL
VYSDSKQFF*

>G361 (54..647)

TCTGTCTCTCTCTCTCTCTTTGTAAATATACATATATAGATAAGCTCACATATATGGCGA
CTGAAACATCTTCTTTGAAGCTCTTCGGTATAAACCTACTTGAAACGACGTCGGTTCAA
ACAGTCATCGGAACCAAGACCCGGATCCGGATCAGGATCCGAGTCACGTAAGTACGAGT
GTCAATACTGTTGTAGAGAGTTTGCTAACTCTCAAGCTCTTGGTGGTCACCAAAACGCTC
ACAAGAAAGAGCGTCAGCTTCTTAAACGTGCACAGATGTTAGCTACTCGTGGTTTGCCAC
GTCATCATAATTTTACCCTCATACCAATCCGCTTCTCTCCGCCCTTCGCGCCGCTGCCTC
ACCTCTCTCTCAGCCGCATCCTCCGCCGCATATGATGCTCTCTCCTTCTTCTCGAGTT
CTAAGTGGCTTTACGGTGAACACATGTCGTCAAAAACGCCGTTGGGTACTTTCATGGTG
GAAGGGGACTTTACGGAGGTGGCATGGAGTCTATGGCCGGAGAAAGTAAAGACTCATGGTG
GTTCTTTGCCGGAGATGAGGAGGTTCCGCCGAGATAGTGATCGGAGTAGCGGAATTAAGT
TAGAGAATGGTATTGGGCTGGACCTCCATTTAAGCCTTGGGCCATGAATGATTATAATTT
TGGCCCAGTAAAGATCTGTAAATACTACTAGGATTTTATTTTATAGAGTATGTTTTTT
TCCTTAATTTTCGGTTGAAATTGGTGAATATTTTATCTCTTACTTACCAATCTCATATT
TCTATGTATGCGTTTGCTTTCACTTTTTTTTATATAATTCTTCTTGTAATAATGCA
ATGTGAGTTTTCTTCCCTATCATCTGTCAAGCTTTGGTTCAATTATTTAGTAATCGAAT
AATATAGGAATAGTGTTGAAAG

>G361 Amino Acid Sequence (domain in AA coordinates: 43-63)
MATETSSLKLFGINLLETTSVQNQSSEPRPGSGSGSESRYECQYCCREFANSQALGGHQ
NAHKKERQLLKRAQMLATRGLPRHNFHPHTNPLLSAFAPLPHLLSQPHPPPHMMLSPSS
SSSKWLYGEHMSSQNAVGYFHGGRGLYGGGMESMAGEVKTHGGSLEPMRRFAGDSRSG
IKLENGIGLDLHLSLGP*

>G562 (137..1285)

ATTTGAATTTCTGGGTTTCTCTCTGTTAAGCTTCTTCTTCTCATCTTCTGCTTACGTT
TCTTCTTCAAGGAGCTTTTCGATTCTTGTAGAAAGAGTCATTGTTCTTCTGAGTGGGAAA
CCTTGAAACCATTTCTATGGGAAATAGCAGCGAGGAACCAAGCCTCTACCAATCAGA
TAAACCATCTTCACCCCGGTGGATCAAACAAATGTTTATGCTTACCTGATTGGGCAGC
TATGCAGGCATATTATGGTCCAAGAGTAGCAATGCCTCCTTATTACAATTCAGCTATGGC
TGCATCTGGTCATCTCTCTCTCTTACATGTGGAATCCTCAGCATATGATGTACCATC
TGGAGCACCTTATGCTGCTGTTTATCCTCATGGAGGAGGAGTTTACGCTCATCCGGTAT
TCCCATGGGATCACTGCCTCAAGGTCAAAGGATCCACCTTTAACAACCTCCGGGACGCT
TTTGAGCATCGACACTCTACTAAATCTACAGGGAACACAGACAATGGATTGATGAAGAA
GCTGAAAGAGTTTGTAGGGCTTGCTATGTCTCTAGGAAATGGGAATCCTGAAATGGTGC
AGATGAACATAAACGATCACGGAACAGCTCAGAACTGATGGTTCTACTGATGGAAGTGA
TGGGAATACAACCTGGGGCAGATGAACCGAACTTAAAAGAAAGTCGAGAGGGAACCTCAAC
AAAAGATGGGAAACAATTGGTTCAAGCTAGCTCATTTTATTCTGTTTCTCCGTCAAGTGG
TGATACCGCGTAAACTCATTCAAGGATCTGGAGCTATACTCTCTCTGGTGAAGTGC
AAATTCACCCCTTCATGTACAATCTTTAGCCATGGTTCTCTGAACTTGGCTTCA

GAACGAGAGAGAACTGAAACGGGAGCGAAGGAAACAGTCTAATAGAGAATCTGCTAGAAG
GTCAAGATTAAGGAAACAGGCCGAGACAGAAGAACTTGCTAGGAAAGTGAAGCCTTGAC
AGCCGAAAACATGGCATTAAAGATCTGAACATAACCACTTAATGAGAAATCTGATAAACT
AAGAGGAGCAAATGCAACCTTGTGGACAACTGAAATGCTCGGAACCCGAAAAGAGAGT
CCCCGCAAATATGTTGTCTAGAGTTAAGAACTCAGGAGCTGGAGATAAGAACAGAACCA
AGGAGACAATGATTCTAACTCTACAAGCAAATTCATCAACTGCTCGATACGAAGCCTCG
AGCTAAAGCAGTAGCTGCAGGCTGAATCGATGGTAATTCATGTCGATTTCTACTTAATTT
GTCGACATAAAACAAAGAAAATAAGTGTACTAATTTTCAGAAAACTTGATAGATAGATAG
TATAGTAGAGAGAGAGAGAGAGAGAGAGGTGTGATGATTATTGATCTATAAATTTTCGGA
GAGAGAGAGGGAGAAAGAGAACTTTTCTCCAGATGAAAATTTGGTGTATGGTTTGT
ACTGTTAATATAGAGAGGCTTTTCTTTTATATAAAATGGCTTCCTTTGTTGCA
>G562 Amino Acid Sequence (domain in AA coordinates: 253-315)
MGNSSSEPKPPTKSKDPSSPPVDQTNVHVYPDWAAMQAYYGPRVAMPPYNSAMAASGHP
PPPYMWNPPQHMMSPSGAPYAAVYPHGGGVYAHPIPMGSLPQGQKDPPLTTPGTLLSIDT
PTKSTGNTDNLMLKLEFDGLAMSLGNGNPENGADHKSRSNSSETDGSDDGSDGNTTG
ADEPKLKRREGTPTKDGKQLVQASSFHSVSPSSGDTGVKLIQSGAILSPGVSANSNPF
MSQSLAMVPPETWLNQERELKRERRKQSNRESARRSRLRKQAEELARKVEALTAENMA
LRSELNQLNEKSKDLRGANATLLDKLKCSEPEKRVANMLSRVKNAGAGDKNKNQGDND
NSTSKFHQLLDTKPRAKAVAAG*
>G591 (88..1020)
GTAAATCTCTCTTTGAAGGTTCTTAACCTCGTTAATCGTAACCTCACAGTGACTCGTTGAG
TCAAAGTCTCTGTCTTTAGCTCAAACCATGGCTAGTAACAACCCCTCACGACAACCTTTCT
GACCAAACCTCTTCTGATGATTTCTTCGAGCAAATCCTCGGCCCTTCTAACTTCTCAGCC
TCTTCTGCCGCCGTTTATCTGGAGTTGACGGAGGATTAGGTGGTGGAGCACCGCCTATG
ATGCTGCAGTTGGGTTCCGGAGAAGAAGGAAGTCAATGGGTGGCTTAGGAGGAAGTGA
CCAACCTGGGTTTCAACATCAGATGTTTCTCTTTGGGGTTAAGTCTTGATCAAGGGAAGGA
CCTGGGTTTCTTAGACCTGAAGGAGGACATGGAAGTGGGAAAAGATTCTCAGATGATGTT
GTTGATAATCGATGTCTTCTATGAAACCTGTTTTCACGGGCAGCCTATGCAACAGCCA
CCTCCATCGGCCCCACATCAGCCTACTTCAATCCGTCCAGGGTTTCGAGCTAGGCGTGGT
CAGGCTACTGATCCACATAGCATCGCTGAGCGGCTACGTAGAGAAAGAATAGCAGAACGG
ATCAGGGCGCTGCAGGAACCTGTACCTACTGTGAACAAGACCGATAGAGCTGCTATGATC
GATGAGATTGTGATTTATGTAAAGTTTCTCAGGCTCCAAGTCAAGGTTTTGAGCATGAAC
CGACTTGGTGGAGCCGGTGCAGTTGCTCCACTTGTTACTGATATGCCTCTTTCATCATCA
GTTGAGGATGAAACGGGTGAGGGTGAAGGACTCCGCAACCGGTGGGAGAAATGGTCT
AACGATGGGACTGAACGTCAAGTGGCTAAACTGATGGAAGAGAAGCTGGAGCCCGCATG
CAGTTCTTCAATCAAAAGGCTCTTTGTATGATGCCAATCTCATTGGCAATGGCAATTTAC
CATTCTCAACCTCCGGATACATCTTCAGTGGTCAAGCCTGAGAACAATCTCCACAGTAG
GATTTCTGCAATAAAGAGTTTGTACAGCTAATCCAACCTGTCCAACATGGGTTTTTCTTCT
GCTCTAATGACTCTGGTTTCTTCTCTCTCTCTCACCAGCTTGAAAGGTAAAAAAGTGAA
AAAGGCTTTGTAGATGGAATCAATGTAGGATTTGCAGTAGAGGGCAAAAAATGTCATAT
AGCTCAATTGATCAAGTCTTAAAAA
>G591 Amino Acid Sequence (domain in AA coordinates: 143-240)
MASNNPHDNLSDQTPSDDFFEQILGLPNFSASSAAGLSGVDGGLGGGAPPMMQLGSGEE
GSHMGLGGSGPTGFHNQMFPLGLSLDQKGPGLRPEGGHGSGKRFSDVDVNRCSMK
PVFHGQPMQPPPSAPHQPTSIRPRVRARRGQATDPHSIAERLRERRERIAERIRALQELVP
TVNKTDRAMIDEIVDYVKFLRLQVKVLSMNRLGGAGAVAPLVTDMPPLSSSVEDETGE
RTPQPAWEKWSNDGTERQVAKLMEENVGAAMQLLOSKALCMMPISLAMAIYHSQPPDTSS
VVKPENPPQ*
>G8 (247..1596)
AAAAAAAATATCCGTCTCACTCTCTCGCCGCCGGTAACATTTCCCGGCGACAAAACCTTC
TCTACTCTCACCATTCTCCATCGTAATCTCTAAATCTTCTCATTCTCTTCTCTCTCC
CGATCATCTCGAGCTCTTCGTGAGAGATTATGTGATTATGTAATCGTTGTTGCTGTAGAA
GACGATCTTAACAACCTGATTTCTTCATCATCACCTTCGCTAGATTGTAATTTTCAGAG
CTTGAGATGTTGGATCTTAACCTCAACGCTGATTCTCCCGAGTCGACTCAGTACGGTGGT
GACTCATACTTAGATCGGCAGACATCAGACAACCTCCGCCGGGAATCGAGTGAAGAGTCC
GGTACATCGACGTCGTCAATTCAATGCCGATGGAGACGAAGACTCTTGCTCTACTCGA
GCTTTCACTCTCAGTTTCGATATTTTAAAGTCGGAAGTAGTAGCGCGGAGACGAAAGC

MLDLNLNADSPSTQYGGDSYLDQRTSDNSAGNRVEESGTSTSSVINADGDEDSCTR
 TFSFDILKVGSSSGGDESPAASASVTKEFFPVSGDCGHLRDVEGSSSRNWIDLSFDRIG
 DGETKLVTVPVTPAPVPAQVKKSRRGPRSRSSQYRGVTFYRRTGRWESHIWDCGKQVYLG
 GFDTAHAAARAYDRAAIKFRGVDADINFTLGDIYEDMKQVQNLKSKEEFVHILRRQSTGFS
 RGSSKYRGVTLHKCGRWEARMGQFLGKKAYDKAAINTNGREAVTNFEMSSYQNEINSESN
 NSEIDLNLGISLSTGNAPKQNGRLFHFPSTYETQRGVSLRIDNEYMGPVNTPLPYGSS
 DHRLYWNGACSPYNNPAEGRATEKRSSEAGMSSNNGWQWRPGQTSAVRFPQPPGPQPPPLFS
 VAAASSGFSHFRPQPPNDNATRGYFYFHP*

GATTTGTCAATTTTTTGTCTAGCCAAAAAAAAAAAAAAAAAGGAGAGAGAGAGAGAGAGA
GAGAGAGAGAGAAACGAAGAAAAAAAAAGAAGCAAAAACATTGTGGGTCTCCGGTGATT
AGGATCAAATTAGGGCACCAGCCTTATCGGAGGAAGAAGCCATGGGTAGAAAAAAGTCG
AGATCAAGCGAATCGAGAACAAAAGTAGTCGACAAGTCAC'TTCTCCAAACGACGCAATG
GTCTCATCGAGAAAGCTCGACAAC'TTCAATTCTCTGTGAATCTTCCATCGCTGTTCTCG
TCGTCTCCGGCTCCGGAAAAC'TACAAGTCTGCCTCCGGTGACAACATGTCAAAGATCA
TTGATCGTTACGAAATACATCATCTGATGAAC'TGAAGCCTTAGACTTTCGCAGAAAAA
CTCGGAATTATCTGCCACTCAAGAGT'TACTAGAAATAGTCCAAAGCAAGCTTGAAGAAT
CAAATGTCGATAATGCAAGTGTGGATACTTAA'TTTCTCTCGGAGGAACAGCTTCGAGACTG
CTCTGTCCGTAAC'TAGAGCTAGGAAGACAGAACTAATGATGGGGGAAGTGAAGTCCCTTC
AAAAAACGGGAGAAC'TTGCTGAGAGAAGAGAAACGAGACTTTGGCTAGCCAGGTGGGGAAGA
AGACGTTTCTGGTTATAGAAGGTGACAGAGGAATGTCTATGGGAAAATGGCTCCGGCAACA
AAGTACGGGAGACTCTTCCGCTGCTCAAGTAATCACCATCATCAACGGCTGAGCTTTCAC
CTTAAACTTACAGCCTGATT'CAGAAGTTTTTACAAATTTGTAAATTATAAAAAAGCTTCAT
AATAATCTCAACCTTTTTATCTTCCTCGCGCCAATGTGGAAATTAAGGTTAAAAATAAAA
TAAACCGAGAAAGCTATCGGAAAGAATTGTAAAACTAAGATAAAGCTATAGTAGATCTTTA
TTGTACCTTCGTAGACGATATAAGATTATTCTGTGTGTTTGTCTTCCCCTCNAAAAAAAA
AAAAAAAAAAAAAAAAA

>G878 (197..1738)

283

AAATATCTTCTCTTTTTTCTGTGTGAGTTGGGTTTGTAAAGTTTATCCTTTTTTGTTC
TCAAAATCAAGAATCGATGGCGGAGAAGGAAGAAAAGAACCATCGAAGTTAAATCATC
CACCGGAGTTTACGGCCAACGATTTCACTACCTCCTCGACCGTTTGGTGAAATGTTTTT
TAGCGGTGGCGTTGGATTTAGTCCTGGACCAATGACTCTCGTCTCAAATTTATTCTCTGA
TCCTGATGAGTTCAAGTCTTTCTCTCAGCTTTTAGCTGGAGCTATGGCTTCTCCGGCGGC
AGCTGCTGTTGCCGCCGCTGCTGTGGTTGCTACTGCTCATCATCAGACACCTGTGAGCTC
TGTCGGTGATGGCGGTGGAAGCGGTGGTGATGTTGACCCGAGGTTTAAAGCAGAGTAGACC
AACGGGATTGATGATAACTCAACCACCGGGGATGTTTACTGTACCGCCGGGGTTAAGTCC
GGCTACTCTTTTGGATTCTCCGAGCTTCTTTGGTCTTTTTTACCTCTTCAGGGAACATT
TGGTATGACACATCAACAAGCTTTAGCACAAAGTCACTGCACAAGCAGTTCAAGGCAATAA
TGTTTCATATGCAGCAATCACAACAATCTGAATATCCTTCTTCTACACAACAACAACA
ACAACAACAACAAGCTTCATTGACTGAGATTCATCATTTTTCTTCTGCACCTAGGTCTCA
GATTCGAGCCTCGGTTCAAGAAACATCGCAGGGTCAGAGAGAGACTTCGGAAATATCTGT
CTTTGAGCATCGGTACAGCCTCAAATGCTGACAAACCAGCTGATGATGGATACAACTG
GCGGAAATATGGGCAGAACCAAGTGAAGGGGAGCGATTTTCTCGGAGTTATTACAAATG
TACGCATCCAGCTTGCTCTGTCAAGAAGAAAGTGGAGAGGTCACTCGATGGACAAGTAAC
GGAAATCATCTACAAGGTCACACAATCATGAGCTTCTCAAAGCGCGGTAACAATAA
CGGGAGTTGTAAAGTTCTGATATTGCAAAATCAGTTTCAAACAAGTAATAGCAGTCTCAA
CAAGAGTAAGAGGGACCGGAAACAAGCCAAAGTTACAACAACAGAGCAGATGCTGTAAGC
AAGTGATAGCGAGGAGGTTGGGAATGCAGAGACTAGTGTTGGGAGAAAGACATGAGGATGA
GCCTGATCCCAAGCGAAGAAATACAGAAGTTCGGGTTTTCAGAACAGTTGCTTCATCGCA
TAGAACTGTGACAGAGCCTAGGATTATTGTCCAAACGACGAGTGAAGTTGACCTCTTAGA
TGATGGATATAGGTGGCGCAAGTATGGTCAGAAAGTAGTCAAAGGAAATCCTTATCCGAG
GAGCTACTATAAGTGTACAACACCAGATTGCGGAGTAAGGAAACATGTAGAGAGAGCAGC
AACTGACCCAAAAGCTGTTGTAACAACATATGAAGGTAAACATAACCATGATGTTCCAGC
TGCTAGAACCAGCAGCCATCAGTTAAGACCAACAATCAACACAACACCTCAACGGTTAA
CTTCAATCATCAACAGCCTGTTGCACGTTTAAAGGCTTAAAGAAGAGCAAATCACTTGACA
GAGAAGAAGAATACGACGCGCTTGAGCTTTTGTGAGTTTAAATGAATCTTCTTTTTTGGTT
AATGAACCTGTTTTTGTGCTCAAACACCACAGGTTTCTCTGGACAGAATCTCTGATA
TTACAGTTTCAAAGGTATGTTCTTTTATTTCATGTTGGAATCTTCTGTGAATCTTAAG
AAGCTTTAGGAGGTAAATGTAAGAAACCAGATTCAAAGTTATGCCCTTATGTGAATCTTT
TGACATGGGATAAACAATAATTTACAGGTATCCTTTTTTGTCTTGTGTAAGAAAAA
AAAA

>G878 Amino Acid Sequence (domain in AA coordinates:250-305, 415-475)

MAEKKEEKEPSKLSSTGVSRPTISLPPRPFGEMFFSGGVGFSPGPMTLVSNLFSDFDEFK
SFSQLLAGAMASAAAAVAAAATAHHQTPVSSVGDGGSGGDVDPFRKQSRPTGLMI
TQPPGMFTVPPLSPATLLDSPSFFGLFSPLQGTFGMTHQQLAQVTAQAVQGNVHMQQ
SQQSEYPSSTQQQQQQQASLTEIPSFSSAPRSQIRASVQETSQGQRETSEISVFEHRS
QPQNADKPADDGYNWRKYGQKQVKGSDFPRSYYKCTHPACPVKKKVERSLDGQVTEIIYK
GQHNHELPPQKRNNGSCSKSSDIANQFQTSNSSLNKSQRDQETSQVTTTEQMSBASDSEB
VGNAETSVGERHEDEPDPKRRNTEVRVSEPVASSHRTVTEPRIIVQTTSEVDLLDDGYRW
RKYGQKVVKGNPYPRSYKCTTPDCGVRKHVERAATDPKAVVTTYEGKHNHDVPAARTSS
HQLRPNNQHNTSTVNFNHQPPVARLRLKEEQIT*

>G971 (131..1171)

TTTTTTTTCTTCCCTCTTTTAGAACTCTCTCTCTCTCTCGTTTTTGACACTTATCCTCTC
TCTTTTTTCTCTCTCCCTCTCTCTCTGCGCGGAAAAAGAACAACGTCGTTTATAGCTAA
AGATTGATCATGTTGGATCTTAACCTAAAGATCTTTTCTTCTTATAACGAAGATCAAGA
TCGGAAAGTACCAATTAATGATCTCAACCACCGGTGAAGAAGAAATCTAATCTATCTCTC
CTCCACAACAGACTCTGCAGCGAGAGATGCTTTCATCGCTTTTGGAAATCTCAAACCGCA
CGATGACCTTGTTCTCTCTCTCTCTCTCTCATAAAGAAACAGGAGATCTCTTTCC
GGTGGTGGCTGATGCTCGTCGGAATATAGAATTCTCCGTGGAAGACAGTCACTGGTTGAA
TCTTTCTTCTTTACAAGAAATACACAGAAAATGGTGAAGAAGAGCAGAAGAGGACCAAG
GTCTCGTAGCTCCCAATATCGTGGCGTCACTTTTTACCGTCCGACCGGTCGTTGGGAATC
TCATATTTGGGATTGTGGAAGCAAGTTTATTTGGGCGGGTTTGATACTGCTTACGCAGC
AGCAAGGGCTTACGACCGAGCTGCTATCAAATTCCGTGGTCTCGATGCAGACATCAATTT
CGTCTGGATGATTATAGGCATGACATCGATAAGATGAAGAATTTAAATAAGGTGGAGTT
CGTGCAACACTTAGCGGAGAGAGTGCGAGTTTCGGAAGAGGAAGTTCCAAATACAAAGG

CTTGGCTCTTCAAAAATGCACCCAATTCAAACTCATGATCAGATTCATCTCTTCCAAAA
CAGGGGATGGGATGCAGCAGCAATAAAATACAATGAGTTGGGAAAGGGAGAAGGAGCCAT
GAAGTTTGGTGCCCATATCAAAGGAAATGGTCACAATGATCTTGAAC TAAGTCTCGGAAT
TTCATCATCATCGGAAAGTATAAAGTTGACAACAGGCGATTACTATAAGGGTATCAATCG
GTCCACGATGGGTTTATACGGTAAGCAATCATCGATATTTTACCCATGGCAACCATGAA
ACCTCTGAAGACAGTTGCAGCATCATCAGGATTCCCTTTTATCAGCATGACAAGTTCCTC
TTCCTCCATGTCCAATTGTTTGTATCCATAGGATCGTTCTACACTCTCTTAATAATATA
TATTTTACTCTATCTGATTATTGTATACAAGGATAAAATTTGATTCTTTCCTTAATGAG
TGAGAAATATTGGAAGTGTAAAAA

>G971 Amino Acid Sequence (conserved domain in aa coordinates: 120-186)

MLDLNLKIFSSYNEDQDRKVLPMISTTGEESNSSSSSTTDSAADAFIAFGILKRDDDL
VPPPPPPPHKETGDLFPVADARRNIEFSVEDSHWLNLSLQNTQKMKVKSRRGPRSR
SQYRGVTFYRRTRWESHIWDCGKVYLGFDYAYAAARAYDRAAIKFRGLDADINFVVD
DYRHDIDKMKNLNKVEFVQTLRRESASFGRSSSKYKGLALQKCTQPKTHDQIHLFQNRGW
DAAAIKYNELGKGEGAMKFGAHIKGNHNDLELSLGISSSES IKLTG DYKGINRSTM
GLYGQSSIFLPMATMKPLKTVASSGF PFISMTSSSSSMSNCFDP*

>G975 (58..657)

ATTACTCATCATCAAGTTCCTACTTTCTCTCTGACAAACATCACAGAGTAAGTAAGAATG
GTACAGACGAAGAAGTTCAGAGGTGTCAGGCAACGCCATTGGGGTTCTTGGGTCGCTGAG
ATTCTCATCTCTCTTGAACGGAGGATTGGCTAGGGACGTTTCGAGACCGCAGAGGAG
GCAGCAAGAGCATACGACGAGGCCCGCGTTTAAATGAGCGGCCGCAACGCCAAACCAAC
TTTCCCTCAACAACAACAACACCGGAGAACTTCCGAGGGCAAAACCGATATTTAGCT
TCGTCCACAATGTATCTCTCAACATCATCTTCATCGCTCTCTTCCATCTCAGCGCCAAA
CTGAGGAAATGCTGCAAGTCTCTTCCCATCCCTCACCTGCCCTCCGTCTTGACACAGCC
AGCTCCCATATCGGCGTCTGGCAGAAACGGGCCGCTTCAAAGTCTGACTCCAGCTGGGTC
ATGACGGTGGAGCTAGGTCCCGCAAGCTCCTCCCAAGAGACTACTAGTAAAGCTTCACAA
GACGCTATTCTTGCTCCGACCACTGAAGTTGAAATTTGGTGGCAGCAGAGAAGAAGTATTG
GATGAGGAAGAAAAGGTTGCTTTGCAAATGATAGAGGAGCTTCTCAATACAACTAAATC
TTATTGCTTATATATATATGTAACCTATTTTCATTGCTGATTACAGCCAAATAATCAATT
ATACCGTGATTTTATAGATGTTTATATTAAGGTTGTTAGATATA

>G975 Amino Acid Sequence (domain in AA coordinates: 4-71)

MVQTKKFRGVRQRHWGSWVAEIRHPLLKRRIWLGTFFETAEEAARAYDEAAVLMSGRNAKT
NFPLNNNTGETSEGKTDISASSTMSSSTSSSLSSILSAKLRKCKSPSPSLTCLRLDT
ASSHIGVWQKRAGSKSDSSWVMTVELGPASSSQETTSKASQDAILAPTTEVEIGGSREEV
LDEEEKVALQMI EELLNTN*

>G994 (180..917)

TGTATATATAGTTAGTTAGTTAGATAAACTTGGTTACCACTTTTGTGTGGTCTTTCTTT
TTCTTTTCTCCATTTTCCATTTATCGACCCCTTGGGTGTAGCTAATTACTTTTCGCGATT
TTCAAATCCAATAAAGTTTAAATTTGATGAAGCTTTTTTAAACCATATAATATAAATAA
TGGGTGGTCGTAAACCATGTTGTGATGAGGTTGGATTAAGAAAGGGTCCATGGACAGTGG
AAGAAGATGGGAAACTAGTTGATTTCTTAAGGGCACGTGGCAACTGCGGTGGTGGTGGAG
GAGGATGGTGTGGAGAGACGTGCCAAACTGGCGGGGCTAAGGAGGTGTGGCAAAAGTT
GCCGTCTCCGGTGGACTAATTATCTCCGGCCAGATCTCAAGAGAGGTCTTTTACTGAAG
AAGAAATCCAAC TAGTCATTGATCTTCATGCTCGCCTTGGCAATAGATGGTCTGAAGATTG
CAGTGGAGTTACCAGGAAGAACAGACAACGATATCAAAAATTATGGAACTCATATAA
AGAGGAAGCTTATAAGAAATGGGTATTGATCCAAACACACATCGTCGATTTGACCAACAAA
AAGTCAACGAGGAGGAAACGATATTGGTCAACGATCCAAAGCCTCTGTCTGAGACCGAGG
TATCTGTTGCTTTGAAGAATGACACGTCAGCAGTGTATCAGGAAATCTAAACCAATTGG
CTGACGTGGACGGTGTATGATCAGCCGTGGAGCTTTCTAATGGAATGACGAAGGAGGAG
GTGGCGACGCCGCCGAGAGCTTACGATGCTATTGTCCGGTGACATTACGTATCATGTT
CTTCTTCGTATCTTTGTGGATGAAGTATGGAGAATTCGGATACGAAGATTTAGAACTTG
GATGTTTCGATGTTTAGAGATTCAAGTATGTTAATTAGCCGTAGGTTGATTAATCATA
AGGTTTATTGACTTATTCTAGAAATGTGTAGTTGGACAGTATAAAGAATCAAAGTTAT
GAAACATTGTAATTTGATTTCCAAATTAATCTAATGAATAAATGTGCTTTGCAAAAAA
AAAAA

>G994 Amino Acid Sequence (domain in AA coordinates: 14-123)

MGGRKPCCDEVGLRGPWTVEEDGKLVDFLRARGNCGGGGGWCWRDVPKLAGLRRCGKS

CRLRWNTNLYLRPDLKRGLFTEEEIQLVIDLHARLGNRWSKIAVELPGRSTDNDIKNYWNTHI
KRKLIRMGIDPNTHRRFDQQKVNEEETILVNDPKPLSETEVSVALKNDTSAVLSGNLNQL
ADVDGDDQPWSFLMENDEGGGDAAGELTMLLSGDITSSCSSSSSLWMKYGEFGYEDLEL
GCFDV*

>G2347 (81..626)

AGCCCATCCTTCAACATTGCTTCCTAACCAGAAATCCACCATCATCTTCCCACGAATACA
ACTTAAAGCTTTACCAGAAAAATGGAGGGTCAGAGAACACAACGCCGGGGTTACTTGAAAG
ACAAGGCTACAGTCTCCAACCTTGTTGAAGAAGAAATGGAGAATGGCATGGATGGAGAAG
AGGAGGATGGAGGAGACGAAGACAAAAGGAAGAAGGTGATGGAAAGAGTTAGAGGTCCTA
GCACTGACCGTGTTCCATCGCGACTGTGCCAGGTCGATAGGTGCACTGTTAATTTGACTG
AGGCCAAGCAGTATTACCGCAGACACAGAGTATGTGAAGTACATGCAAAGGCATCTGCTG
CGACTGTTGTCAGGGGTCAGGCAACGCTTTTGTCAACAATGCAGCAGGTTTCATGAGCTAC
CAGAGTTTGATGAAGCTAAAAGAAGCTGCAGGAGGCGCTTAGCTGGACACAATGAGAGGA
GGAGGAAGATCTCTGGTGACAGTTTTGGAGAAGGGTCAGGCCGGAGAGGGTTTAGCGGTC
AACTGATCCAGACTCAAGAAAGAAACAGGGTAGACAGGAACTTCCTATGACCAACTCAT
CAATCAAGCGACCACAGATCAGATAAACCCCTCCCGTCTCTCTCTTCTGTCATCTACATA
TGCTCTATCTACACTCTTATAGACAAATAATGGCATCTAACAATGTCAAGAAAAGTTGG
TCATGGTATTAAATCCTACACGGATATATAACTATAAACCTCTAGTCCCCTCTATGCTGT
CCTGTAATGAATATCTATCCGAAATGTATTCGCATAGTCTTGCGTCTAATAATGTTTAT
TGATTTTGTA

>G2347 Amino Acid Sequence (domain in AA coordinates: 60-136)
MEGQRTQRRGYLKDKATVSNLVEEEMENGMDGEEEDGDEDEKRRKVMERVGRPSTDRVPS
RLCQVDRCTVNLTEAKQYYRRHRVCEVHAKASAATVAGVRQRFCCQCSRFFHELPEFDEAK
RSCRRRLAGHNERRRKISGDSFGEBSGRRGFSGQLIQTERNRVDRKLPMTNSSFKRPQI
R*

>G2010 (1..525)

ATGGAGGGTAAGAGATCACAAGGACAAGGTTACATGAAAAAGAAGTCTTACCTTGTGGAA
GAAGATATGGAGACTGATACGGATGAAGAAGAGGAAGTAGGTAGGGATAGAGTTAGAGGG
TCTAGAGGTAGCATCAATCGTGGTGGCTCGTTGCGGCTTTGCCAAGTAGATAGATGCACA
GCTGATATGAAAGAGGCAAAACCTGTATCACCGGAGACACAAAGTGTGTGAAGTTCATGCA
AAGGCATCTTCTGTCTTTCTCTCAGGACTTAACCAACGCTTTTGTCAACAATGCAGTAGG
TTTCATGACCTCCAAGAGTTTGATGAAGCTAAGAGAAGTTGCAGGAGGCGCTTAGCTGGA
CACAATGAGCGAAGAAGGAAGAGCTCTGGTGAGAGTACTTATGGAGAAGGATCAGGTCGG
AGAGGAATCAATGGTCAGGTGGTGTATGCAGAAATCAAGAAAGATCAAGGGTAGAGATGACA
CTTCCTATGCCAAACTCATCATTCAAGCGACCACAGATTAGATAG

>G2010 Amino Acid Sequence (domain in AA coordinates: 53-127)
MEGKRSQGGQGYMKKSYLVEEDMETDTDEEEVGRDRVRGSRGSINRGSLRLCQVDRCT
ADMKEAKLYHRRHKVCEVHAKASSVFLSGLNQRFCQCSRFDLQEFDEAKRSCRRRLAG
HNERRRKSSGESTYGEBSGRRGINGQVVMQNQERSRVEMTLMPNSSFPRPQIR*